

Estadística aplicada a las Ciencias Sociales

Curso destinado a la preparación de futuros estudiantes de las asignaturas de Estadística de grados universitarios en CC. Sociales

Ramón Pérez Juste

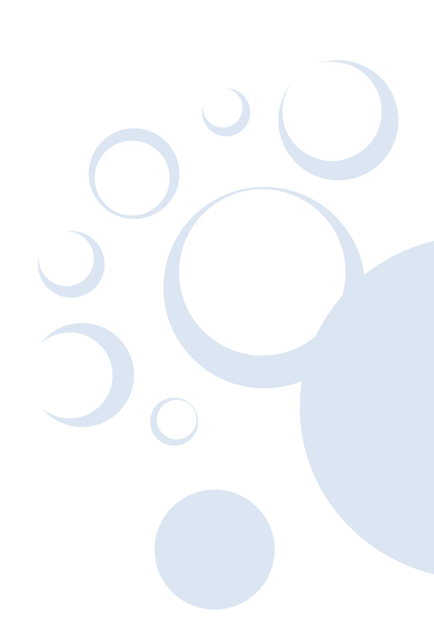


1ª Edición: Madrid, noviembre de 2012

Índice

Segunda parte: Pruebas de autoevaluación

Tercera parte: Glosario



PRIMERA PARTE
Contenidos fundamentales

PRESENTACIÓN DEL CURSO

Les sugiero que accedan a la siguiente presentación:

http://www.intecca.uned.es/portalavip/grabacion.php?ID_Grabacion=60989&ID_Sala=65271&hashData=37ddf4107dda75b332b5e43ad57e11fe&paramsToCheck=SURfR3JhYmFjaW9uLElEX1NhbGEs

Para muchos alumnos, la asignatura de Estadística resulta ser una de las más difíciles. En ocasiones, la dificultad de la misma es intrínseca, es decir, se encuentra ligada a sus objetivos, contenidos y niveles de exigencia; sin embargo, en otras, deriva de una previa actitud de "respeto" hacia los números -material con el que se trabaja en la asignatura- probablemente relacionada con carencias en la formación previa, y a la falta de sentido que muchos estudiantes le atribuyen en cuanto a las aportaciones para su formación académica y profesional.

Ante esta situación, con el curso cero pretendemos servir de preparación para abordar el estudio de los contenidos de la asignatura con mayores probabilidades de éxito; junto a ello, y en íntima relación, buscamos modificar la actitud con la que muchos alumnos se enfrentan a su estudio a través de una doble contribución:

a) Poner de relieve la utilidad y valor de los conocimientos que aporta para su formación académica y profesional.
b) Hacerle ver que es capaz de alcanzar los objetivos y competencias de las diversas asignaturas estadísticas.

En torno a la primera cuestión, valga el siguiente testimonio de un alumno del curso 2011-12, expresado al finalizar el mismo:

Al comenzar el estudio de esta asignatura, creía que no tenía nada que ver con la educación social, pero al estudiarla me he dado cuenta de lo importante que es tener una base en los métodos de investigación, ya que cuando ejerzamos de educadores sociales antes de aplicar nada habrá que hacer diagnóstico y para ello la estadística nos ayuda a tomar la decisión correcta, y saber elegir los instrumentos más adecuados para ello. Los métodos de investigación es una ciencia muy amplia… rica para aportar soluciones con las que cualquier profesional estaría mucho más organizado, tranquilo, seguro, teniendo conocimientos de esta disciplina. Con la realización de este trabajo he aprendido y he comprendido mucho más, que en el estudio anterior de la asignatura. Realizando casos se aprende mejor la estadística. Es decir, se aprende mejor con la práctica que con la teoría y yo he aprendido mucho de esta asignatura.

O este otro de una alumna del mismo curso:

He aprobado Estadística gracias al trabajo y quisiera saber donde lo puedo ver y la nota del mismo.

A pesar de lo complicado que me resultó hacerlo, mi emoción aumentaba con cada punto resuelto. Gracias a él he podido entender algo más la Estadística y me gustaría saber los errores que he cometido para poder subsanarlos junto con la corrección de los mismos, si es posible.
Por último, agradecerles la flexibilidad que nos han brindado a causa de la dificultad de la materia al tener en cuenta el trabajo realizado.

Pues bien: lo deseable es iniciar el estudio de la asignatura con una actitud semejante, en lugar de llegar a ella después de la experiencia más o menos prolongada de estudio por pura obligación.

Por ello, el enfoque del curso no se centra en el cultivo de destrezas de cálculo y utilización de fórmulas, sino en la *comprensión* de sus procedimientos, procesos y aportaciones, obviando cuanto sea posible el estudio teórico. Y digo "cuanto sea posible" porque, como alguien ha dicho, la mejor práctica es una buena teoría.

ORGANIZACIÓN DEL CURSO

El curso consta de tres partes diferentes; la presente, primera, se centra en la explicación y desarrollo de los contenidos fundamentales. La segunda se dedica a las pruebas de autoevaluación; por último, el curso se completa con un Glosario, que permite acceder a los conceptos fundamentales marcados en el presente documento mediante asterisco (*).

En relación con los contenidos, de las dos grandes modalidades de Estadística, la *descriptiva* y la *inferencial,* el curso se centra en la primera, dando a conocer los principales análisis y tratamientos de datos para una y dos variables*.

En el primer caso se presentan y analizan las medidas más utilizadas: además de las individuales, como las *puntuaciones directas*, diferenciales* y típicas* o los cuantiles*, las de grupo, tales como las de *posición*, dispersión* y forma**; en el segundo nos acercaremos a los conceptos de *correlación** y de *regresión*, junto a alguna técnica específica de correlación*, orientada a la comprensión del concepto.

Tanto en uno como en otro caso, los conceptos obtenidos a partir de números se presentan, además, mediante representaciones gráficas, orientadas, de nuevo, a facilitar la comprensión de los conceptos y las consecuencias que se pueden seguir de los análisis de datos, en particular, las posibles decisiones para la práctica.

Por otra parte, con la finalidad de ofrecer un curso con la menor cantidad de teoría posible, como ya hemos indicado acudiremos a marcar con asterisco aquellos términos que aparecen desarrollados en el Glosario.

Digamos, por último, que, para comprender sus aportaciones más relevantes, se harán algunas referencias a la *estadística inferencial* y, por consiguiente, a la *teoría de la probabilidad** ya que es en ese contexto donde la estadística nos ofrece una mayor utilidad y aportaciones. Lo más importante en este apartado será el acercamiento al estudio de un modelo* estadístico fundamental: *la curva normal de probabilidades*.*

CAPÍTULO 1. APORTACIONES DE LA ESTADÍSTICA

Como hemos señalado, es importante que el lector llegue a apreciar el estudio de la asignatura por sus aportaciones, primero a su formación académica y, como consecuencia, a su capacitación para el futuro profesional.

Creo que a ello puede contribuir escuchar la siguiente grabación:

http://www.intecca.uned.es/portalavip/grabacion.php?ID_Grabacion=61012&ID_Sala=65319&hashData=5043c7f6e56448fd82ca9a0d38e7fb25&amp%3BparamsToCheck=SURfR3JhYmFjaW9uLElEX1NhbGEs

La Estadística le capacita para:

- ***Interpretar las puntuaciones individuales*** de los sujetos en el contexto de los grupos de los que forman parte.*
- ***Caracterizar los grupos** con los que trabaje: una clase, un curso, los miembros de una profesión (médicos, abogados, fontaneros, jornaleros…), los sujetos que han realizado una encuesta, etc.*
- ***Extraer información** de tales características para la toma de decisiones de carácter profesional.*
- ***Identificar relaciones*** existentes entre las puntuaciones obtenidas por los miembros de esos grupos en dos o más variables*.*
- ***Aplicar** a los miembros de los grupos las propiedades de los modelos* estadísticos a los que se acomodan los datos empíricos.*
- ***Poner a prueba** diferentes formas de intervención –métodos, formas de motivación, sistemas de disciplina, castigos, premios…- sobre sujetos o grupos.*

Desarrollemos brevemente estos aspectos:

1. Interpretar puntuaciones individuales

- A todos nos resulta fácil interpretar la talla o el peso de una persona como elevada, media o baja. Así, una puntuación individual*, por ejemplo 80 Kg y 168 cm, nos son relativamente fáciles de interpretar en nuestro contexto de referencia. De una manera más o menos precisa, nos hacemos idea de cómo se encuentra esa persona en relación con los miembros del grupo del que forma parte.
- Pero esta interpretación puede ser más precisa si conocemos determinadas características del grupo, tales como la "normalidad" de talla y peso en el grupo de edad o sexo del que forma parte. A esto nos ayuda la Estadística, indicándonos cuál es la media aritmética* del grupo y en cuánto se aparta de esa puntuación* el sujeto de que se trate (dispersión*).
- Sin embargo, esto mismo no es tan fácil si hablamos de variables* diferentes, como la estabilidad emocional, la autoestima, la inteligencia, la sociabilidad, el nivel de conocimientos… La posibilidad de "medir" estas variables, con las limitaciones a que haremos referencia más adelante, nos pone en una situación próxima, aunque no tan exacta como la anterior.
- Por otra parte, medidas específicas, como los cuartiles, la edad mental, el cociente intelectual o las puntuaciones típicas nos permitirán una interpretación más técnica y precisa.

2. Caracterizar grupos

- Como acabamos de decir, la interpretación de las puntuaciones individuales suele hacerse en el contexto de los grupos de los que esa persona forma parte. Las ideas políticas de un individuo en relación con la clase social a la que pertenece, la inteligencia en el contexto de su edad y sexo, las calificaciones de Inglés en el seno de la clase de 4º de ESO, los niveles de colesterol, la talla o el peso, teniendo en cuenta la edad, el sexo o el grupo de riesgo de que forma parte… son ejemplos de datos que debemos ser capaces de interpretar y valorar.
- Pues bien: la Estadística, mediante las medidas de *posición*, dispersión* y forma**, nos informa de esas características grupales.
- Con estos valores, quienes deban tomar decisiones o hacer interpretaciones especializadas (profesores, orientadores, psicólogos, sociólogos, economistas…) pueden hacerlo con mayor seguridad de acertar que desconociendo tales datos.

3. Extraer información de tales características para la toma de decisiones

- Los científicos, los estudiosos, los profesionales y, en general, las personas interesadas en los diferentes campos del saber, no acuden a la Estadística por sí misma sino por la utilidad que les proporciona la información que les ofrece.
- Un profesional de la Educación encontrará información relevante para organizar las actividades en su aula, para atender a la diversidad de sus alumnos, para mejorar sus programas, para predecir (y tomar decisiones preventivas) sobre los alumnos con riesgo…
- Un psicólogo podrá caracterizar a sus pacientes, diagnosticar síndromes, recomendar tratamientos…
- Un sociólogo será capaz de orientar a los políticos, interpretar estados sociales, adelantarse a las crisis…
- Un economista ayudará a la empresa a prevenir problemas, a identificar riesgos, a diseñar campañas atendiendo a los perfiles de los clientes…
- Un médico podrá estar al tanto de la incidencia de ciertas enfermedades, de los riesgos de determinados medicamentos, de las peculiaridades de ciertos pacientes en relación con algunos fármacos…
- En fin: la utilidad de la Estadística tiene que ver con su ayuda a los profesionales para tomar decisiones que les son propias.
- Y todo lo anterior –dimensión práctica- no reduce sus aportaciones al puro avance del saber, interés primordial del científico en sus diferentes ámbitos del conocimiento, sino que lo engrandece.

4. Identificar las relaciones* entre variables

- El ser humano, como persona aislada o formando parte de grupos, se comporta de modos muy diversos como consecuencia de la interacción entre sus características y las del contexto en que vive y de las relaciones existentes entre unas y otros.
- La identificación de las relaciones entre variables de la persona o de estas con características individuales o grupales aporta gran información al sociólogo, al economista, al psicólogo o al pedagogo, incluso al médico. En fin: es una rica información para los profesionales que trabajan con personas.

- Pues bien: la Estadística nos permite conocer si ciertas variables están relacionadas con otras o no, esto es: si varían conjuntamente (co-varían) o son unas independientes de otras. Por ejemplo, si la inteligencia está o no relacionada con la clase social, si la autoestima se relaciona con la introversión, si agresividad y seguridad en sí mismo son independientes o están relacionadas…En estadística, representamos por lo general la correlación* con el símbolo r_{xy}, esto es: la correlación entre las variables X e Y (por ejemplo: entre inteligencia y rendimiento académico, entre pobreza y analfabetismo…)
- Esta información es fundamental para identificar las variables sobre las que poder intervenir cuando se desea modificar –positiva o negativamente- otra variable. Por ejemplo, conocer las variables que están ligadas (relacionadas) con la autoestima, nos ayuda a intervenir sobre aquellas para modificar esta. Sabiendo cómo se relaciona la motivación con el rendimiento, podemos incidir sobre aquella para elevar este; conocer la relación entre el dinero en circulación y el grado de inflación ayuda al político a tomar las medidas pertinentes, etc.
- La Estadística nos informa sobre estas variables, sobre el tipo de relación (positiva: elevando los valores de una se elevan los de la otra, y viceversa) o negativa (elevando los valores de una disminuyen los de la otra y al contrario) y sobre su intensidad (perfecta, imperfecta –lo habitual, más o menos elevada- o nula).
- Es más: a través de ciertas propiedades de las correlaciones podemos *predecir*, bien es verdad que con márgenes de error y determinados niveles de probabilidad*, lo que ocurrirá en una variable conociendo los valores obtenidos en otra. Por ejemplo: un orientador puede predecir, asumiendo cierto riesgo de equivocarse, qué alumnos suspenderán al final de curso en Estadística, a partir de una variable relacionada con ella, como son ciertas competencias matemáticas. Lógicamente, la intervención irá destinada a evitar que se cumpla la predicción. El margen de error ocurre en toda predicción, como la del tiempo, de la evolución de una enfermedad, de la famosa "prima de riesgo", de las actitudes hacia los extranjeros, etc.
- En la figura 1 podemos apreciar de forma intuitiva cómo un predictor como la inteligencia mantiene una relación con el éxito académico de aproximadamente 0,60, lo que viene a representar que explica poco más de una tercera parte del criterio (zona rayada de la figura 1.a).

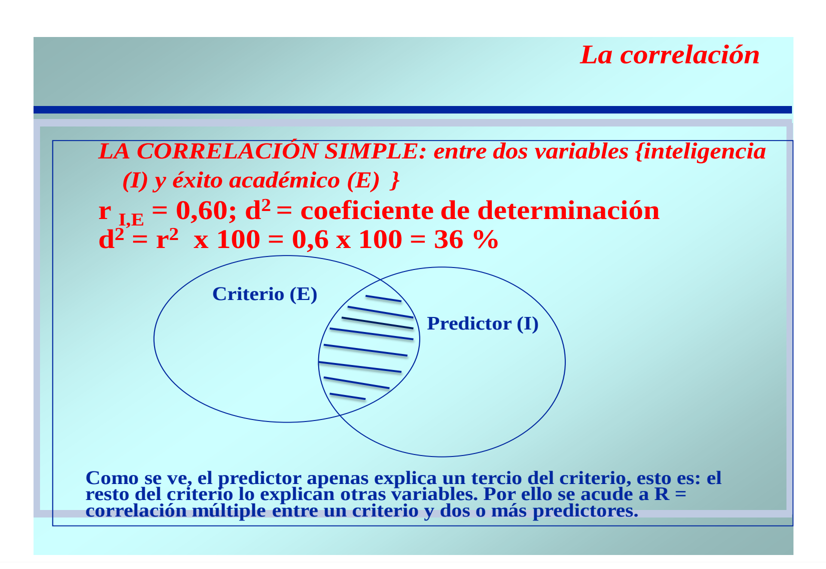


Figura 1 a. Rrepresentación intuitiva de la correlación simple entre dos variables

- La Estadística aborda este problema con la correlación* múltiple (simbolizada por $R_{1.234...n}$), donde un mismo criterio se intenta predecir con varios predictores. Una

aproximación intuitiva se presenta en la figura 1.b. No obstante, este tema no es objeto de trabajo en nuestro curso.

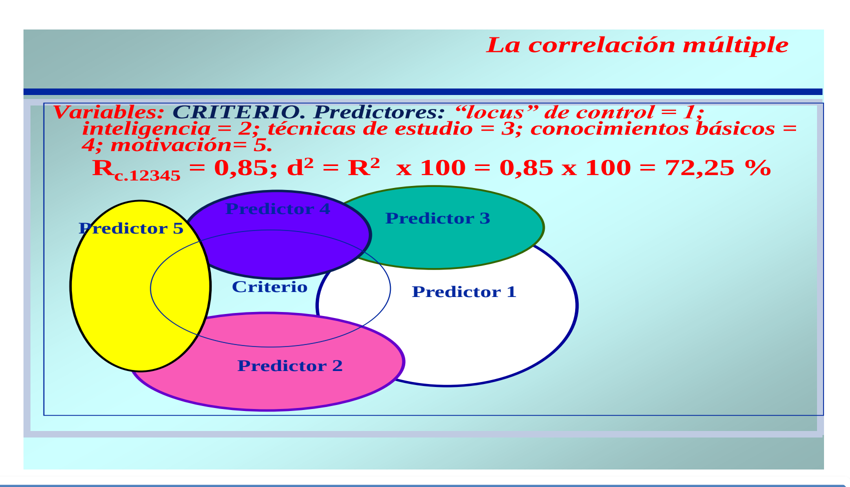


Figura 1 b . Representación intuitiva de la correlación múltiple R

5. Aplicación de las propiedades de los modelos estadísticos

Creo que todo lo anterior ya justifica el estudio y dominio de los contenidos de la asignatura. Sin embargo, y aunque lo que viene a continuación exige ya una cierta base, las principales utilidades de la Estadística, están ligadas a su modalidad *inferencial* a la que solo haremos una breves referencias.

En esencia, esta parte de la Estadística trata de ir más allá de los datos empíricos, datos obtenidos mediante instrumentos como los test, los exámenes, los cuestionarios, las encuestas, la observación, las entrevistas… Como hemos anunciado, con ellos hemos podido interpretar una puntuación individual*, caracterizar un grupo o averiguar si se dan o no relaciones entre variables. Ahora la cuestión es más compleja: con los valores *medidos* a los integrantes de esos grupos, ¿podemos ir más allá y utilizarlos para interpretar las puntuaciones de otros sujetos que forman parte de grupos con las mismas características?

Veamos:

Lo normal es que en Estadística trabajemos con los valores obtenidos por un grupo de sujetos de una edad, sexo, curso, carrera, raza, clase social, ideología, religión, … en variables como actitudes, conocimientos, técnicas de estudio, autoconcepto, *locus of control*, nivel de pobreza… pero el interés del investigador es que, a partir de ellos, se puedan aplicar –con cierta prudencia y admitiendo márgenes de error- al conjunto de sujetos de la misma edad, sexo, curso, carrera, en la variable estudiada.

Los primeros valores, denominados *estadísticos**, se "miden" en conjuntos denominados *muestras*;* los segundos son *estimados* para el conjunto total, denominado *población*;* los valores *estimados* se denominan *parámetros*.* Estimar es tanto como atribuirle un valor mediante procedimientos técnicos; no obstante, cualquier estimación está sujeta a errores que deben ser tomados en consideración, como lo hace la Estadística (*error de estimación).*

Pues bien: para poder hacer tal cosa, la Estadística aplica a los datos muestrales las propiedades de ciertos modelos*, para lo cual lo primero es decidir si a aquellos se les puede aplicar el modelo y sus propiedades.

Podemos entender esto fácilmente. No creo que nadie haya visto jamás en la realidad un cono perfecto; sin embargo, todos identificamos los volcanes –pensemos en el Teide- con una forma cónica. Admitiendo que la realidad nos presenta objetos cónicos –más o menos cercanos al cono ideal- podemos aplicar a tales objetos reales las propiedades del cono; del mismo modo podríamos actuar con el prisma, con la pirámide, con la esfera…, y calcular así la superficie y el volumen de un objeto piramidal, prismático o esférico.

Un caso similar y sencillo en nuestro ámbito; todos conocen o han oído hablar de la denominada curva normal de probabilidades* o campana de Gauss. En sí misma es un modelo*, por tanto, algo ideal: no encontrarán en la naturaleza ninguna realidad igual a esa campana, pero sí datos más o menos próximos a ella. Pues bien: lo que se plantea es que si los datos reales se aproximan *razonablemente bien* al modelo –y esta es ya una cuestión estadística- sus propiedades puedan aplicarse a los datos reales, lo que supone un gran avance en el tratamiento de la información recogida.

En las figuras siguientes (Fig. 2, a, b y c) podrán apreciar el modelo normal –en el centro- y dos series de datos empíricos, más o menos cercanos al mismo. Decidir si se les pueden aplicar las propiedades de modelo normal es la cuestión que nos ayuda a resolver la Estadística:

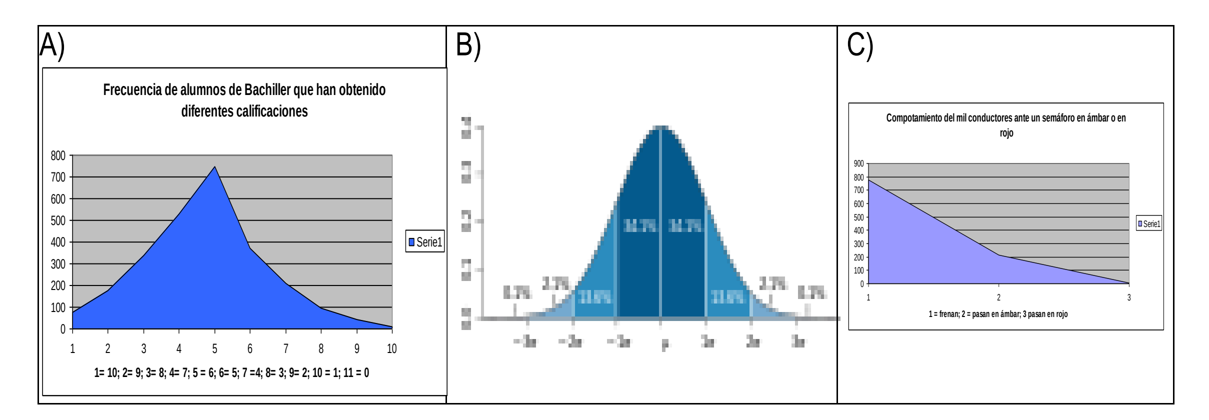


Figura 2 a, b y c: Representaciones de tres series de datos, A, B y C

Tomando como normal la figura central, ideal, modelo teórico, parece claro que la primera figura se acerca más a ella que la tercera. Podemos descartar que esta sea “normal”, pero no tenemos seguridad de que la primera sí lo sea. La Estadística nos ayudará a decidirlo asignando a la decisión una determinada probabilidad* de estar en lo cierto. En caso afirmativo, podremos aplicarle las propiedades del modelo, al igual que utilizamos las del cono para estimar el volumen de una montaña cónica.

Obviamente, al aplicar tales propiedades somos conscientes de ciertas deformaciones; pero también es cierto que estas pueden deberse a que nuestra muestra* no era lo suficientemente representativa del conjunto de la población* por problemas de tamaño y de forma de seleccionar sus componentes. En ese caso, esas deformaciones afectan a los datos empíricos pero estos serían más y más cercanos al modelo en la medida en que corrigiéramos tales deformaciones.

Esto nos tranquiliza y nos permite sentirnos autorizados para aplicar las propiedades del modelo a los datos empíricos.

6. Poner a prueba diferentes formas de intervención

La Educación supone siempre una intervención sobre personas o grupos con ánimo de modificarlas de forma perfectiva. ¿Qué intervención, de entre varias posibles, es la más eficaz en el logro de los objetivos que persigue?. Por ejemplo: ¿cuál de entre varios métodos de enseñanza, o de motivación, o de modificación de conducta… es más eficaz para alcanzar los objetivos?

La Estadística nos ayuda a poner a prueba en condiciones adecuadas –mediante *diseños experimentales**- una o más formas de intervención –variables independientes*- y a contrastar sus efectos contra la posibilidad de que no sean eficaces o de que no podamos asegurar que los efectos apreciados se deben a ellas sino a otras causas que, en general, definimos como *azar**.

Es evidente que podemos *medir* los efectos de la variable independiente* en unos determinados grupos de estudiantes; pero tan evidente como esto puede ser que al profesor le interese que lo que ha constatado en los grupos objeto de investigación valga para otros grupos con los que comparta características; en definitiva: que pueda aplicar a la *población** los resultados obtenidos en las *muestras**. Y esto es *inferencia*. El caso más claro será que lo que ha comprobado mediante la investigación con los alumnos de un curso académico lo pueda aplicar a los del siguiente curso y a cursos posteriores.

Creo que todos estos aspectos, aportaciones importantes, deben hacer al lector comprender que el estudio de la Estadística* le merece la pena más allá del puro aprobar la asignatura y que, en consecuencia, se ponga a la tarea pensando, por una parte, que le será de utilidad y, por otra, que puede no ya aprobarla sino formarse para su futura profesión. A ello les animo encarecidamente.

PARA SU REFLEXIÓN

¿Con qué actitud se enfrenta usted al estudio de la Estadística?

¿Considera que tiene carencias de base en Matemáticas básicas que pueden dificultar el dominio de la asignatura?

¿Considera que las aportaciones que le ofrece la Estadística, que le acabamos de presentar, merecen un estudio de la misma?

A la vista de las aportaciones reseñadas, ¿considera que su actitud inicial hacia la asignatura ha mejorado? Si no es así, ¿a qué lo achaca?

¿Considera que el anterior contenido le ha sido de utilidad?

A su juicio, comprender y ser capaz de utilizar los conocimientos del tema, ¿le serán de utilidad cuando comience a estudiar la asignatura?

A su juicio, comprender y ser capaz de utilizar los conocimientos del tema, ¿le serán de utilidad en su vida profesional?

Es probable que haya cuestiones del tema que no haya comprendido. Le invitamos a comunicarlas en el cuestionario de evaluación del curso.

CAPÍTULO 2. LA ESTADISTICA TRABAJA CON NÚMEROS

Es habitual oír a estudiantes que ellos son de “letras” y que los números siempre les han supuesto una gran dificultad y hasta cierta aversión.

Sin embargo, los números están en la vida ordinaria de las personas, en toda su actividad diaria. Por tanto, hay que utilizarlos con soltura y para ello, nada mejor que comprender, desde la base, desde lo más elemental, lo que representan, sus grandes aportaciones y los análisis y tratamientos que se hacen con los mismos.

Para ello, para empezar, les recomiendo acceder a la siguiente presentación:
http://www.intecca.uned.es/portalavip/grabacion.php?ID_Grabacion=61053&ID_Sala=65333&hashData=3c90c6e5ff1f62a9f3745bb298a43c00&amp%3BparamsToCheck=SURfR3JhYmFjaW9uLElEX1NhbGEs

1. ¿De dónde salen los números?

Imagine a un investigador en Psicología, Sociología, Pedagogía, Ciencias Económicas, Medicina… en su despacho. Ha recogido los protocolos de un cuestionario de opinión sobre sus respectivos objetos de estudio: actitudes hacia los inmigrantes, opiniones políticas, motivación, niveles de renta, incidencia de un virus… que se acumulan sobre su mesa de trabajo. Puede tener 100, 200, 1000 o más cuestionarios.

Para que se haga una idea, en el cuadro 1 se presenta parte de un cuestionario utilizado en una reciente investigación (puede encontrarla en
http://www.fundacionabaco.org/index.php?modo=descargas&m=84&idcat=79

Cuestionario para la mejora de la Educación en Andalucía

A continuación encontrará una serie de enunciados relacionados con la calidad de la educación, seguidos de un espacio destinado a recoger el grado de importancia que, a su juicio, reciben ***de hecho*** en el centro educativo en el que presta sus servicios. Su valoración se expresa entre 1, la mínima, y 4, la máxima. En la siguiente columna deberá marcar si, a su juicio, es ***manifiestamente necesario*** introducir mejoras en tal aspecto.

ITEMS	Importancia que se le concede en su centro:				Necesidad de mejora	
	1 -	2	3	4 +	SI	NO
I. UN PROYECTO EDUCATIVO DEL CENTRO:						
1. Que incorpore los valores morales y sociales de consenso						
2. Que integre el carácter propio en el caso de los centros concertados y privados						

3. *Que analice y tome en consideración las necesidades, demandas y expectativas de la sociedad actual y, en particular, de su propio contexto*						
4. *Que tenga en cuenta las características de nuestro tiempo: globalización, migraciones, interculturalidad, nuevas tecnologías…*						
5. *Que sea conocido por las familias*						
6. *Que sea asumido por las familias, comprometidas con sus valores*						
7. *Que cuente con un sistema de revisión periódica y de puesta al día*						

Cuadro 1: Parte de un cuestionario utilizado en el Informe Ábaco

¿Cómo pasamos de esos protocolos a los números con los que trabaja la Estadística*?

Esos cuestionarios pueden contener 10, 20, 50… preguntas o *ítems* que los sujetos pueden valorar, por ejemplo, asignando un 1 si responde SI, y un 0 cuando marca el NO. Sumando los valores marcados llegamos a una puntuación* con el número de "síes" y de "noes". Del mismo modo, dado que es posible marcar entre 1 y 4 la valoración de cada ítem, según la importancia concedida a cada uno de los enunciados, podremos obtener la puntuación* total de cada persona consultada y hasta la tendencia del grupo, merced a una medida tan conocida como es la media aritmética.*

Al final, cada uno de tales protocolos se ha convertido en un número, número con el que trabajará la Estadística. Este número se denomina *puntuación directa** y suele representarse por X_i, esto es: puntuación directa o bruta del sujeto *i.*

Recuerde: la puntuación directa* de un sujeto cualquiera -también llamada "bruta"- en un instrumento de recogida de datos, se representa por X_i y se lee: puntuación directa del sujeto i.

Los números que manejamos nacen de pesar, medir o contar los "objetos" más diversos. Objetos como la talla, el peso, la edad, la inteligencia, la asertividad, la renta "per cápita", la autoestima, el rendimiento académico, la masa muscular, el "ranking" de un país en los Juegos Olímpicos (o en las pruebas PISA)… son objetos de medida y, mediante los instrumentos adecuados, dan lugar a números.

1.1. Diferentes tipos de números

Parece claro, no obstante, que, por encima de la apariencia externa de los números -todos son iguales en su apariencia- en realidad son muy diferentes unos de otros. Los 80 cm. de talla de un niño poco tienen que ver con los 80 puntos en una prueba de inglés, o los 80 Kg de peso, o el puesto 80 al llegar a meta, o el número de sujetos -80- que son admitidos a un concurso, o los 80º centígrados alcanzados por un horno, o los 80 puntos obtenidos en una prueba de autoestima, o…

Cuando hablamos de talla, o de peso, estamos ante los números plenos, además de fiables y válidos si han sido medidos con cuidado y utilizando un metro y una balanza fiables. Y ello se debe a que contamos con unidades de la misma naturaleza que el objeto a medir: el centímetro para la talla o el kg. para el peso, y damos por hecho que quien mide lo hace con seriedad.

Aparentemente, los 80º centígrados del horno son de la misma naturaleza que los anteriores, pero no es así, por una sencilla razón: si antes el 0 significaba que no tenemos delante a nadie (porque nadie pesa 0 gr. ni mide 0 cm.), ahora, como todos sabemos, por debajo de 0º sigue habiendo temperatura: -3º, -15º, etc.

También se puede asignar ese número a un corredor de maratón que ha llegado a meta en el puesto 80. Pero aquí no medimos la distancia recorrida sino el *orden* de llegada; y puede haber diferencias notables en minutos o segundos entre las llegadas de los diferentes atletas. Puede ocurrir que entre el primero y el segundo apenas haya un par de segundos pero que entre este y el tercero haya más de un minuto, y que, más adelante, entre un grupo de corredores prácticamente en el mismo tiempo pero distanciados del anterior en 15 o 20 minutos. Pero eso no importa si lo que medimos es el *orden* en que entran.

Y, obviamente, el número de 80 de admitidos a un determinado concurso, 35 varones y 45 mujeres por ejemplo, solo nos indica las veces que personas varones y personas mujeres han sido seleccionadas, sin más.
Un mismo valor numérico puede representar objetos reales o empíricos muy diferentes; según sean estos objetos, al número que los representa se les podrán aplicar unas u otras propiedades de los números y sus correspondientes operaciones matemáticas.

1.2. Escalas de medida

Pues bien: cada uno de esos 80 representa un tipo diferente de número, propios de diferentes *niveles o escalas de medida*: los de *razón* o *cociente*, en el primer caso (talla, peso), permiten todo tipo de operaciones aritméticas; los de *intervalo*, en el segundo (grados centígrados), con los que podemos establecer ciertas operaciones pero no otras (no conviene entrar aquí en detalles); los de *orden* (puesto ocupado al llegar a meta) nos indican solo lo que es mayor o menor, anterior o posterior, más o menos intenso…, pero no podemos operar con ellos de otra manera; o de tipo *nominal,* que solo nos indican que algo es igual o diferente que otro algo, pero no podemos hacer operaciones con ellos: asignar un 1 a los varones y un 2 a las mujeres no significa que estas sean más –o aquellos, menos- que los varones, sino, simplemente, *diferentes*. No tendría sentido, por lo tanto, sumar el número de unos y el doses e intentar calcular la media.

Ahora bien: observe el lector algo importante: hay objetos fácilmente medibles, porque están abiertos a nuestros sentidos (talla, peso, edad…) y porque tenemos unidades de medida de la misma naturaleza: cm, gr., año…

Pero hay otros que son, en realidad, objetos cuya misma naturaleza no conocemos y, por ello, tenemos que *definirlos* previamente. Pensemos en la inteligencia, en la asertividad, en la autoestima, en la opinión… y hasta en el rendimiento académico. Nadie ha visto la inteligencia, pero sí a personas inteligentes, o asertivas, o con baja autoestima, o con rendimiento adecuado…

Para “medirlas” debemos, en primer lugar, *definirlas*. A Binet se debe una famosa frase cuando se le preguntó ¿qué es la inteligencia? Su respuesta fue tan contundente y clara como discutible: Inteligencia –dijo- es lo que mide mi test. Podríamos decir que, a partir de otros autores de tests la inteligencia, llegaríamos a diferentes medidas de este rasgo humano (Y así es, por cierto). Y lo mismo pasará con el rendimiento académico: diferentes exámenes darán lugar a diferentes resultados, diferentes medidas. Y nada digamos si hablamos de asertividad, de autoestima, de

esquizofrenia, etc. Una vez definido el objeto, debemos encontrar manifestaciones acordes con la definición o elaborar reactivos que se consideren evidencias del mismo. Estos reactivos o estas manifestaciones se convierten en *ítems* del instrumento de medida. Por lo general, a esta traducción de la definición a ítems se la llama definición operativa u operacional.

Aquí nos encontramos con serios problemas para encontrar una *regla de medida* y su correspondiente *unidad de medida* y, por tanto, para asignar valores numéricos a la realidad medida. He aquí un problema que tendrán que conocer en su estudio de la asignatura.

Medir determinados objetos de los ámbitos en que trabajamos – Educación, Economía, Medicina, Psicología, Sociología…- implica ***definir*** el objeto a medir, encontrar manifestaciones de tal objeto o reactivos adecuados y decidir la ***regla de medida***, la regla que nos permitirá atribuir un valor a cada manifestación o reactivo ***unidad de medida)***.

Nosotros dejamos constancia de tal problema, señalando las limitaciones que ello representa para los números que utilizamos, en particular:

a) *Para las operaciones matemáticas que están justificadas con tales números.*
b) *Para su fiabilidad: los números obtenidos en una ocasión pueden variar en otra*
c) *Para su validez: podemos estar midiendo una cosa que no es por completo la cosa deseada*.

1.3. Variables y escalas de medida

Ciertos "objetos" no presentan manifestaciones diferentes. Se les denomina *constantes.* Sin embargo, otros si las tienen, tales y se les denomina *variables*;* tal es el caso del sexo, masculino o femenino; del estado civil: soltero, casado, divorciado o viudo; de los grados universitarios: Pedagogía, Psicología, Sociología, Económicas…; junto a estos, en otros casos los objetos a medir admiten valores que difieren en *cantidad*. A las primeras, las denominamos variables *cualitativas**, y se miden con números propios de escalas *nominales* mientras las segundas se conocen como *cuantitativas*,* y admiten números ordinales, de intervalo o de razón. Las cualitativas pueden presentar dos categorías –*dicotómicas**, como en el caso del sexo- o más, en cuyo caso hablamos de cualitativas *politómicas*,* como ocurre con el estado civil.

Algunos autores hablan de variables *cuasi-cuantitativas*,* en las que la cantidad solo puede apreciarse en términos de *orden*, por lo que son propias de escalas ordinales. Una variable de este tipo es la *escala de dureza de los cuerpos* en la que cada cuerpo está por delante o detrás de otro según que le raye o sea rayado por él (A Friedrich Mohs se debe una escala de 10 niveles de dureza que van del talco, el más blando, al diamante, al que solo puede rayar otro diamante).

A su vez, estas variables *cuantitativas* se dividen en *discretas** (variables continuas que solo admiten valores enteros, como número de hijos o de alumnos) y *continuas**, como es la edad, el peso, la talla…donde podemos asignar todos los valores intermedios si disponemos de los instrumentos adecuados (una balanza de precisión, por ejemplo; o un cronómetro, como el utilizado en las pruebas olímpicas de atletismo). En el cuadro 2 aparece una clasificación de las variables.

TIPOS DE VARIABLES	ESCALA DE MEDIDA
Cualitativas	Nominales • Dicotómicas: sexo • Politómicas: estado civil, clase social, grado universitario…
Cuasi-cuantitativas	Ordinales: • Escala de dureza • Rangos o puestos • Clasificación de los terremotos
Cuantitativas	Continuas: edad, talla, peso Discretas: número de alumnos

Cuadro 2: Clasificación de la variables

En algunos de estos casos caben transformaciones; así, una variable cuantitativa *continua* puede ser "tratada" como *discreta*, prescindiendo de algunos de los valores posibles; por ejemplo: podemos tomar valores de edad de los alumnos quedándonos con los años y obviando los meses, o tomando años y meses, obviando semanas, días…

Una dificultad añadida se da en el caso de datos cualitativos, como los surgidos de entrevistas, que se desea tratar tanto cualitativa como cuantitativamente. Por ejemplo, en el citado *Informe Ábaco*, utilizamos el siguiente guión para las entrevistas a personalidades representativas de la sociedad andaluza, tanto en general como del ámbito educativo en particular (Cuadro 3):

Guión para las entrevistas en el "Informe Ábaco"

¿QUÉ PRETENDEMOS?

Recoger información subjetiva de los encuestados que nos permita conocer:

- Una valoración ***global y genérica*** sobre la calidad del sistema educativo andaluz.
- Los aspectos en que fundamenta su valoración (en definitiva: qué entiende por calidad del sistema educativo)
- En qué lugar sitúan el sistema andaluz en el contexto español: por debajo, en la media o por encima
- Los aspectos en que, a su juicio, el sistema andaluz está mejor y peor que la media
- Cuáles son los aspectos, a su juicio, que necesitan una acción de mejora más urgente. Cuáles son, a su juicio, los más difíciles de afrontar
- Las medidas que tomaría en el supuesto de tener plena autoridad, y autoridad efectiva, para mejorar el sistema

PREGUNTAS A FORMULAR

1. ¿Cuál es, a su juicio, la valoración global de la calidad de la educación andaluza? Trate de calificarla como MUY BUENA, BUENA, ACEPTABLE, REGULAR, MALA
2. Cuando ha emitido tal valoración, ¿en qué aspectos concretos ha pensado o ha tenido en cuenta?

3. Si pone en relación la calidad de la educación en Andalucía, en general, ¿dónde la sitúa: por encima, por debajo o en la media?
4. Tal vez haya aspectos en los que la educación en Andalucía pueda situarse de forma diferente en relación con la de España. Según su juicio, ¿hay aspectos en los que la Educación en Andalucía está por encima de la media? ¿Cuáles?. ¿Y por debajo de la media? ¿Cuáles?
5. ¿Cuáles son los aspectos más necesitados de una mejora urgente? ¿Por qué?¿Cuáles son los aspectos más difíciles de mejorar? ¿Por qué?
6. Póngase en el supuesto de que tiene autoridad efectiva para cambiar las cosas a mejor. ¿Qué medidas tomaría en primer lugar?¿Cuáles encontrarían una mayor resistencia? ¿Cuáles exigirían una mayor prudencia?

Cuadro 3: Objetivos y preguntas para la entrevistas del Informe Ábaco

Se comprenderá que el tratamiento de estos datos puede y debe ser muy diferente (no procede aquí sino anunciar técnicas adecuadas mediante programas informáticos, como Atlas-ti o AQUAD). Pero cabe también acudir a un análisis de frecuencias, comprobando las veces que los 88 consultados dan un determinado tipo de respuestas, por ejemplo, qué % de los que responden consideran que la calidad de la educación en Andalucía se sitúa en la media, por encima o por debajo de la media de España (pregunta 3); luego, los consultados, se extenderán en los por qué de su respuesta, que serán analizados cualitativamente, pero ese dato es cuantitativo, de nivel meramente nominal.

De hecho, la transcripción de las 88 entrevistas ocupó más de mil folios, que fueron analizados cualitativamente mediante técnicas específicas.

El tratamiento de datos cualitativos es mucho más complejo que el que se da a instrumentos de recogida de datos en los que los sujetos consultados pueden atribuir valor a sus respuestas de acuerdo con determinadas reglas de medida. No es objeto de este curso 0 abordar este punto.

2. Los números nos informan, los números dicen cosas

Cuando un sociólogo hace una encuesta sobre intención de voto, obtiene determinados valores que suele traducir a porcentajes para su interpretación.

Cuando un psicólogo aplica un test de autoestima a un grupo de alumnos, asigna a cada uno una puntuación*; esta puntuación oscila entre un suelo y un techo (mínima y máxima), cuyos valores dependen de la regla de medida y de su correspondiente unidad de medida (por ejemplo: un punto por cada respuesta positiva).

Cuando un profesor propone a sus alumnos un examen, asigna a cada uno de ellos una calificación que, del mismo modo, depende del número de ítems o preguntas y de la regla de medida: por ejemplo: número de respuestas acertadas menos número de errores, partido por el número de alternativas ofrecidas menos 1, fórmula habitual para la calificación de las pruebas objetivas (ecuación 1):

Puntuación (X_i)

$$Ecuación\ 1:\ \ Xi = \sum_{i=1}^{N} A - \frac{E}{a-1}$$

Obviamente, “medir” el rendimiento con un examen de desarrollo multiplica los problemas para decidir cuál es la unidad y, en consecuencia, cuál es el valor a asignar a cada examen.

Veamos algunos casos:

- Un cuestionario de 40 preguntas (*ítems*) en que el encuestado puede marcar SI, NO, NO SÉ.
- Una prueba objetiva en la que el profesor decide valorar solo las preguntas bien resueltas.
- Otra prueba objetiva en la que el profesor aplica la fórmula anterior, restando los errores teniendo en cuenta que las alternativas ofrecidas son 4.
- Una escala de actitud en que para cada ítem el consultado debe marcar su posición entre 1 (mínimo) y 7 (máximo).

Parece evidente que si en cada una de esas situaciones una persona obtiene 25 puntos, tal valor no puede interpretarse del mismo modo ni puede significar lo mismo.

Debemos tomar conciencia de la importancia que cobra la ***regla de medida***; con ella atribuimos valor –números- a la información recogida con los diferentes instrumentos. Ahora bien: conviene reflexionar sobre el carácter ***arbitrario*** que, en muchas ocasiones, tiene la decisión sobre tal regla, y lo que ello representa para el trabajo con tales números.

2.1. ¿Cómo interpretar esos valores? El caso de las puntuaciones individuales

Si nos interesa una puntuación individual* (representada por X_i: puntuación directa* del sujeto i) solo nos hacemos una idea de dos maneras: conociendo los valores extremos o poniendo su puntuación en relación con el grupo. Por ejemplo, 9 puntos es un sobresaliente si la prueba se puntúa sobre 10, pero es una puntuación muy baja si lo es sobre 50. Y también es una puntuación baja si la mayoría de las puntuaciones obtenidas por los sujetos del grupo están por encima de los 30 puntos.

Otra forma de interpretar las puntuaciones es a través de determinadas transformaciones de las puntuaciones individuales directas, a ciertas medidas, como puede ser un *cuantil.* Entre los cuantiles, los más usados son el *cuartil, el decil y el centil o percentil;* estas medidas nos indican la posición de un sujeto cuando el grupo se ordena en cuatro, diez o cien partes. Así, estar en el *cuartil* 1 (Q_1) es encontrase entre el 25 % inferior del grupo; hallarse en el *decil* 7 (D_7) equivale a superar al 70 % del grupo, y obtener una puntuación equivalente al *centil o percentil* 78 (C_{78}) viene a ser superar al 78 % del grupo.

Medidas individuales son, también, la puntuación de desviación –*puntuación diferencial**- representada por *x*, que no es sino su separación –negativa o positiva- en relación con la media del grupo ($x_i = X_i$ – *Media)*.

Así, un sujeto cuya x_i = -2 nos indica, de entrada, que puntúa por debajo de la media aritmética* (signo negativo) y, en concreto, que se aparta dos puntos de la misma.

La Edad Mental (EM) es, también una puntuación individual*, y su cociente con la edad cronológica (EC) otra diferente, el Cociente Intelectual (CI): *CI = EM / EC.* La EM indica que una persona tiene una inteligencia propia de una determina edad. Por ejemplo, si un niño tiene EM = 9, estamos diciendo que su desarrollo intelectual equivale al de un niño ideal de 9 años; claro está: si tal niño tiene en realidad 12, estamos afirmando que tiene retraso mental, pero si tuviera 8, la interpretación es que es un niño con desarrollo intelectual por encima del propio de su edad. Para una mejor interpretación se ha desarrollado el CI, por lo general multiplicado por 100. De esta forma, un niño de 6 años y EM de 6, tiene un CI = 1, o de 100, si lo multiplicamos por 100. Tanto ese 1 como el 100 nos informan de un niño cuyo desarrollo intelectual es normal, apropiado a su edad cronológica.

Otra medida, que exige previamente el cálculo de medidas de grupo (a las que nos referiremos en seguida), es la puntuación z_i; esta puntuación individual es el cociente entre la *puntuación diferencial* (x_i)* de cada persona (puntuación directa menos la media del grupo) y la desviación típica* de este. En resumen, la z_i indica en cuántas desviaciones típicas del grupo se aparta un sujeto cualquiera de la media del mismo (ecuación 2). Para entendernos: lo mismo que hablando de distancias utilizamos como unidad el Km., en este caso tomamos con unidad el valor de la desviación típica*.

Entenderemos mejor esto al hablar en su momento de la curva normal de probabilidades*. En efecto: con ella presente sabremos que z_i = 0 representa a un sujeto normal, en la media; que z_i = -1 es propia de un sujeto que supera al 34 %, mientras que z_i = + 1 lo hace con el 84 % aproximadamente.Clarificaremos estos conceptos más adelante.

$$Ecuación\ 2: \quad z_i = \frac{x_i}{s} : \frac{(X_i - Media)}{s}$$

Medidas individuales son aquellas que se refieren a un solo sujeto; como se ha indicado, su puntuación directa* se representa por X_i. Para interpretar este valor podemos acudir a x_i o *puntuación diferencial* * con respecto a la media; a z_i, que indica en cuántas desviaciones típicas se aparta el sujeto de la media aritmética* del grupo; o a los diversos cuantiles (Q, D o P). Existen otras medidas, como la EM o el CI.

2.2. El caso de las puntuaciones en grupo

Si nuestro interés es interpretar las puntuaciones del grupo, y este es pequeño, no resulta difícil hacernos una idea de cómo es ese grupo; sin embargo, cuando es grande, ver las puntuaciones tal y como van apareciendo al ser calificados los exámenes o valorado un test, o recogidas las puntuaciones de un cuestionario… se convierte en algo complejo: los números parecen una realidad confusa e informe, como se aprecia en el siguiente conjunto de datos:

Serie de datos 1:

72, 87, 95, 88, 79, 69, 55, 54, 69, 77, 88, 60, 64, 60, 88, 77, 67, 75, 75, 52,
52, 67, 77, 95, 87, 60, 95, 86, 77, 67, 85, 51, 51, 67, 77, 85, 94, 64, 64, 50,

94, 93, 85, 76, 64, 75, 91, 82, 85, 62, 62, 77, 82, 91, 90, 80, 85, 82, 110, 75,
62, 62, 75, 72, 80, 62, 94, 90, 67, 85, 54, 60, 90, 72, 80, 22, 79, 89, 57, 89,
79, 8, 57, 77, 71, 76, 89, 91, 54, 70, 94, 79, 57, 55, 70, 89, 70, 88, 26, 10
N = 100

Ante estos hechos, la Estadística* nos ayuda mediante la organización de los datos, en particular a través de su *ordenación* y *reducción* o *simplificación*

3. Organización de los datos

La primera operación que suele realizarse es la de *ordenar* los números, las puntuaciones. Una operación tan sencilla como esa nos permite conocer:

- *Las puntuaciones extremas; puestas las puntuaciones individuales en relación con las extremas posibles del cuestionario, de la prueba objetiva,… ya nos ofrecen una información interesante.*
- *La continuidad o no de las mismas, apreciando si se dan o no valores vacíos, huecos.*
- *La acumulación o no y en qué parte –superior, central o inferior- de la distribución de valores ordenados.*

Veamos varios casos en una sencilla escala entre 0 y 10:

Caso a): 8, 6, 6, 6, 5, 3, 3, 3, 2, 2.

- Aquí apreciamos que no aparecen las puntuaciones extremas (9 y 10, 0 y 1)
- Que hay valores vacíos: 7 y 4
- Que se da una doble acumulación de puntuaciones, una en la parte superior y otra en la inferior (6 y 3, repetidos en tres ocasiones).

Caso b): 9, 8, 7, 6, 5, 5, 5, 4, 3, 2

En esta ocasión tampoco el grupo presenta puntuaciones a lo largo de todo el recorrido (falta el 10, el 1 y el 0), pero no hay huecos (hay mayor continuidad que en el caso anterior), se da una acumulación en el centro (valor 5) y una notable simetría en torno a la puntuación central.

Caso c)
1, 1, 1, 1, 1, 1, 1, 1, 1, 1
5, 5, 5, 5, 5, 5, 5, 5, 5, 5
9, 9, 9, 9, 9, 9, 9, 9, 9, 9

En estos tres ejemplos la distribución es uniforme: todos los sujetos alcanzan la misma puntuación, pero en el primero todas son bajas y en el tercero todas elevadas, frente a al segundo, de puntuaciones medias.

Comparemos ahora estas dos:

5, 5, 5, 5, 5, 5, 5, 5, 5, 5
10,10, 10, 10, 10, 0, 0, 0, 0, 0

Como vemos, estamos ante la máxima homogeneidad y la máxima heterogeneidad respectivamente. Cuando calculemos las medias aritméticas, veremos que en ambos grupos la media es la misma (5), pero un profesor que tuviera que trabajar con uno u otro grupo debería actuar claramente de formas bien distintas.

Los casos anteriores nos ilustran sobre el valor de una operación tan simple como es la *ordenación* –creciente o decreciente- de las puntuaciones. Fácilmente se comprenderá que esa utilidad es mucho mayor si, en lugar de los 10 casos, tuviéramos ante nosotros 100, 400, 1000…

Pero en ciertos casos, cuando el tamaño es elevado –pongamos 100 o más casos- la ordenación es laboriosa y su utilidad queda limitada, como fácilmente se desprende de los datos anteriores que utilizaremos más adelante.

La forma más sencilla de hacernos cargo de ciertas características de un grupo consiste en ***ordenarlos*** de forma creciente o decreciente. Esta sencilla acción nos permite apreciar su ***recorrido*** (diferencia entre las puntuaciones extremas), si se da o no ***continuidad*** a lo largo del mismo, su ***dispersión* o variabilidad*** o la ***forma**** y el lugar en que se agrupan las puntuaciones.

3.1. La reducción. Distribución de frecuencias.

Cuando el conjunto de casos es elevado, como en la anterior serie 1 (el valor de N es de 100) una forma de facilitar la interpretación es mediante la *reducción* del conjunto a otro menor. El caso más sencillo se da cuando se evita la repetición de las puntuaciones. Estamos hablando de una *distribución de frecuencias* en la que, por un lado, tenemos las puntuaciones obtenidas (X_i) y, por otra, las veces que cada puntuación aparece en el conjunto de casos (f_i).

Así, con los casos a y b del apartado anterior podríamos reducirlos, quedando del siguiente modo (Cuadro 4):

Caso a): 8, 6, 6, 6, 5, 3, 3, 3, 2, 2.
Caso b): 9, 8, 7, 6, 5, 5, 5, 4, 3, 2

Caso a)		Caso b)	
X_i	f_i	X_i	f_i
8	**1**	**9**	**1**
6	**3**	**8**	**1**
5	**1**	**7**	**1**
3	**3**	**6**	**1**
2	**2**	**5**	**3**
		4	**1**
		3	**1**
		2	**1**

Cuadro 4. Reducción de datos originales a una distribución de frecuencias

Supongamos que hemos realizado un examen, consistente en una prueba objetiva, a un total de 30 alumnos. Las calificaciones, a fin de que sean fácilmente comprensibles, las hemos reducido a la escala 0 - 10, habitual en el ámbito académico. Cabe pensar que estas calificaciones se puedan considerar ordinales e, incluso, de intervalo, dado que disponemos de una unidad de medida razonablemente precisa. He aquí los datos (Serie 2):

X_i: 9, 7, 7, 4, 5, 6, 7, 3, 1, 8, 8, 9, 3, 4, 10, 6, 3, 4, 8, 7, 1, 3, 2, 5, 7, 5, 4, 5, 8, 2

Podemos hacer la ordenación de mayor a menor, que ya nos informará de las características de este grupo de alumnos:

X_i: 10, 9, 9, 8, 8, 8, 8, 7, 7,7, 7, 7, 6, 6, 5, 5, 5, 5, 4, 4, 4, 4, 3, 3, 3, 3, 2, 2, 1, 1

Como vemos:

- *Únicamente nos falta la puntuación 0*
- *Se da continuidad de las puntuaciones*
- *Apreciamos una mayor concentración de puntuaciones elevadas*

Si reducimos el conjunto de datos, podemos apreciarlo más claramente. Para ello basta construir una distribución de frecuencias. Nótese que entre la serie anterior y la siguiente no se dan sino diferencias de forma pero no de contenido:

X_i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**N**
f_i	**2**	**2**	**4**	**4**	**4**	**2**	**5**	**4**	**2**	**1**	**30**

Tabla 1: Distribución de frecuencias correspondiente a la serie 2

Como vemos, las 30 puntuaciones han quedado reducidas a 10 diferentes y la acumulación de las puntuaciones repetidas, tomadas como frecuencias, nos permite una mayor y más fácil comprensión de las características del grupo:

a) *Heterogéneo*
b) *Continuo: no se aprecia discontinuidad entre las puntuaciones*
c) *Con tendencia hacia las puntuaciones más elevadas: si tomamos el 5 como suficiente o aprobado, 18 de las 30 lo alcanzan y lo superan.*
d) *Además, en lugar de concentrarse el mayor número de casos en lo que podríamos llamar lo "normal", esto es: en torno al 5, comprobamos que, además de los de puntuación muy alta (el 9 y el 10), lo que predomina son las puntuaciones elevadas (los notables).*

Estas características son de gran relevancia para un profesor que deba atender las diferencias entre sus alumnos, o para un orientador que tenga que trabajar la autoestima de los mismos.

Los datos anteriores pueden presentarse de forma intuitiva mediante una representación gráfica conocida como *histograma** (figura 3) consistente en un eje de coordenadas, con los diferentes valores en el eje de abscisas y con las frecuencias en el de ordenadas.

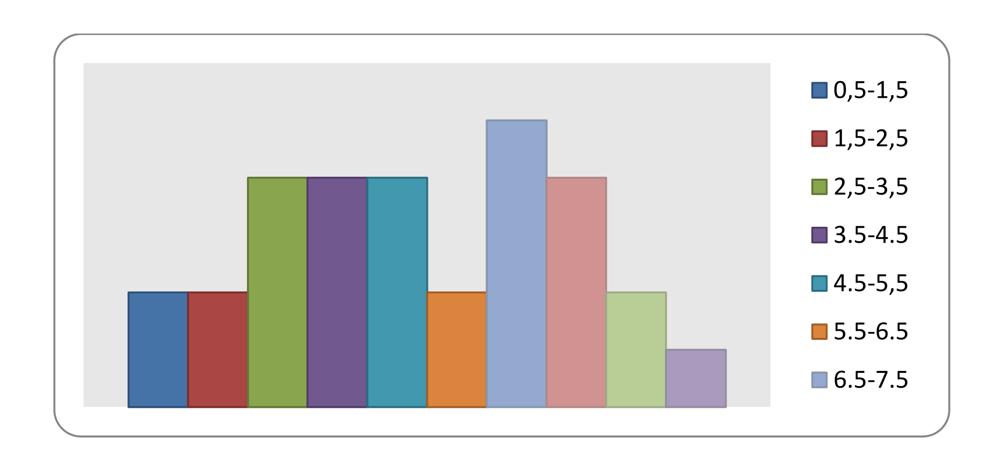


Figura 3. Histograma correspondiente a los datos de la tabla 1

Una distribución original de datos, ordenada o no, puede reducirse por medio de una distribución de frecuencias; en ella se presenta una columna -o una fila- con las diversas puntuaciones, representadas por X_i, y otra con las frecuencias -f_i- o veces que cada puntuación se repite.
La representación gráfica adecuada es el histograma.

Sin embargo, todavía es posible una reducción mayor de los datos, algo necesario cuando el *rango* o *recorrido* de las puntuaciones (diferencia entre los valores extremos, representado por R) es mayor, como ocurre con los siguientes datos, ya aludidos previamente (serie 1), en donde estamos ante 100 sujetos (N = 100) con puntuaciones que pueden ir de 0 a 130, presentados primero de forma "natural", desordenada, y luego ordenados en forma decreciente:

Serie desordenada, con las puntuaciones X_i según aparecen al investigador (Serie 1):

72, 87, 95, 88, 79, 69, 55, 54, 69, 77, 88, 60, 64, 60, 88, 77, 67, 75, 75, 52,
52, 67, 77, 95, 87, 60, 95, 86, 77, 67, 85, 51, 51, 67, 77, 85, 94, 64, 64, 50,
94, 93, 85, 76, 64, 75, 91, 82, 85, 62, 62, 77, 82, 91, 90, 80, 85, 82, 110, 75,
62, 62, 75, 72, 80, 62, 94, 90, 67, 85, 54, 60, 90, 72, 80, 22, 79, 89, 57, 89,
79, 8, 57, 77, 71, 76, 89, 91, 54, 70, 94, 79, 57, 55, 70, 89, 70, 88, 26, 10
N = 100

Serie ordenada en forma decreciente, correspondiente a los datos anteriores:

110, 95, 95, 95, 94, 94, 94, 94, 93, 91, 91, 91, 90, 90, 90, 89, 89, 89, 89, 88,
88, 88, 88, 87, 87, 86, 85, 85, 85, 85, 85, 85, 82, 82, 82, 80, 80, 80, 79, 79,
79, 79, 77, 77, 77, 77, 77, 77, 77, 76, 76, 75, 75, 75, 75, 75, 72, 72, 72, 71,
70, 70, 70, 69, 69, 67, 67, 67, 67, 67, 64, 64, 64, 64, 62, 62, 62, 62, 62, 60,
60, 60, 60, 57, 57, 57, 55, 55, 54, 54, 54, 52, 52, 51, 51, 50, 26, 22, 10, 8.
N = 100

El rango, en este caso es: R = 110 – 8 + 1 = 103 puntuaciones diferentes posibles. La mera ordenación ya nos permite ver el amplio recorrido de las mismas, con valores que, por una parte,

se acercan a las puntuaciones más extremas (8, cerca del 0, y 110, próximo a la puntuación máxima de 130)

Pero, por otra parte, si reducimos la serie a las puntuaciones directas (X_i) con sus correspondientes frecuencias (f_i), como hemos hecho en el caso anterior, podemos apreciar que estamos ante una distribución todavía muy amplia todavía de no fácil apreciación de una forma global e intuitiva: nada menos que 35 valores:

X_i: 110, 95, 94, 93, 91, 90, 89, 88, 87, 86, 85, 82, 80, 79, 77, 76, 75, 72
f_i: 1 3 4 1 3 3 4 4 2 1 6 3 3 4 7 2 5 3

X_i: 71, 70, 69, 67, 64, 62, 60, 57, 55, 54, 52, 51, 50, 26, 22, 10, 8
f_i: 1 3 2 5 4 5 4 3 2 3 2 2 1 1 1 1 1

Tabla 2. Distribucion de frecuencias (Amplitud del intervalo = 1) correspondiente a los datos de la serie 1

Por ello es frecuente que la distribución tome la modalidad de *intervalos*, esto es: se trata de una distribución que nos indica cuantos casos (frecuencias: f_i) hay para un conjunto de puntuaciones que denominamos *intervalos* (I). Lógicamente, las frecuencias serán tanto más elevadas cuanto menor sea el número de intervalos. Por ello hay que decidir con prudencia cuántos intervalos teniendo en cuenta el recorrido del conjunto y la amplitud que queremos dar a cada uno.

A tales efectos, debemos pensar que siempre que hagamos una distribución de intervalos vamos a "deformar" la distribución original en mayor o menor grado, ya que a todas las puntuaciones del intervalo las vamos a representar por una, la que ocupe el lugar central de cada intervalo (*marca de clase*, representada por X_i, al igual que la puntuación directa). Sin embargo, cabe pensar que las deformaciones en un intervalo en un sentido tenderán a compensarse con las de otros intervalos en sentido contrario. Veamos.

En nuestro caso, la distribución oscila entre 8 y 110 puntuaciones; por tanto, hay (110 – 8) + 1 puntuaciones posibles; podemos hacer una distribución por intervalos; si tomamos la decisión de que su amplitud sea de 10 puntos, la distribución podría ser la siguiente (Tabla 3):

I	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-110	
X_i	5.5	15.5	25.5	35.5	45.5	55.5	65.5	75.5	85.5	95.5	105.5	
f_i	2	2	0	0	1	16	19	25	23	11	1	$\sum = 100$
X_if_i	11	31	0	0	45.5	888	1244,5	1887,5	1966,5	1050,5	105.5	$\sum = 7230$

Tabla 3: Distribución de frecuencias (amplitud del intervalo = 10) correspondiente a los datos de la serie 1.

Como se puede apreciar, esta distribución es mucho más manejable y hasta intuitiva; a simple vista apreciamos:

- *Su gran heterogeneidad*

- *Su discontinuidad en la parte inferior, con dos grandes huecos de puntuaciones carentes de sujetos (a partir de la puntuación 20 hasta la 40, ambas inclusive). Cabe pensar que los cuatro sujetos inferiores de los dos primeros intervalos de la distribución podrían considerarse ajenos al grueso de grupo.*

- *La tendencia a valores elevados, no solo por el caso que se encuentra en el intervalo superior sino porque las mayores frecuencias se sitúan claramente a la derecha de la misma*

- *Debemos reconocer cierta distorsión. La más clara está en la puntuación superior, 110, que queda disminuida al ser representada por la marca de clase (105.5), al igual que la inferior, 8, que quedará representada por 5.5. Sin embargo, se acepta que en otros casos ocurrirá al contrario y que, en conjunto se compensan. En efecto, la puntuación 22 será representada por 25,5 y la 94 por 95.5.*

- *Además, no debemos olvidar que, en general, estamos trabajando con números que no son totalmente fiables, que su fiabilidad no es total, por lo que la aparente pérdida de precisión no es tal si reconocemos esas limitaciones de los números que utilizamos.*

Una representación gráfica de estos datos es el *histograma**, con una base proporcional a la amplitud del intervalo y una altura relaciona con su frecuencia (fig. 4a):

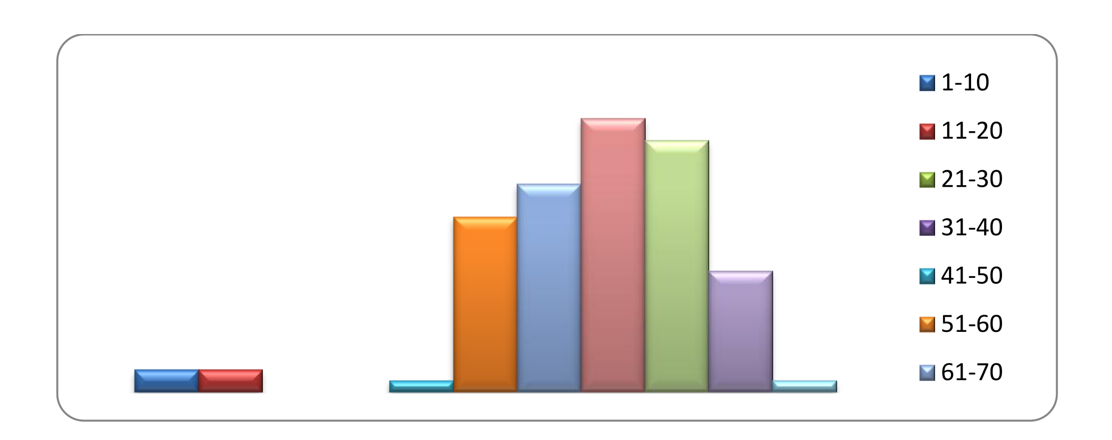


Figura 4a: Histograma correspondiente a los datos de la serie 1

Cuando el recorrido de la variable* es muy amplio, es preferible acudir a una distribución por intervalos. En este caso, utilizamos una fila, o una columna para los intervalos y otra para las frecuencias. En teste caso, las frecuencias son las correspondientes a la amplitud de cada intervalo (que comienza medio punto antes de su puntuación inferior y acaba medio punto después de la puntuación superior). Cuando se opera con este tipo de distribuciones, cada intervalo se representa por su ***marca de clase***, X_i, igual que la puntuación directa en el caso de datos no agrupados.

En el caso en que las frecuencias correspondieran a una distribución de frecuencias de variables cualitativas, como pueden ser los diferentes estados civiles o los grados universitarios, la representación se denomina *diagrama de barras** (figuras 4.b y 4c).

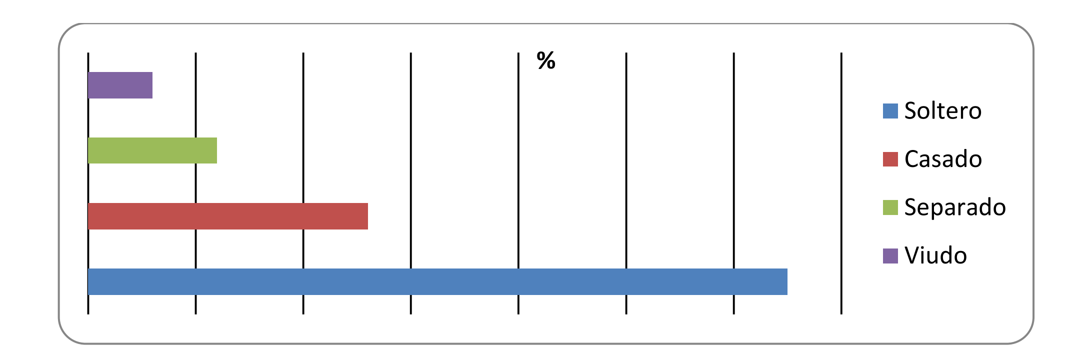


Figura 4b: Diagrama de barras correspondiente a los % de estudiantes según su estado civil

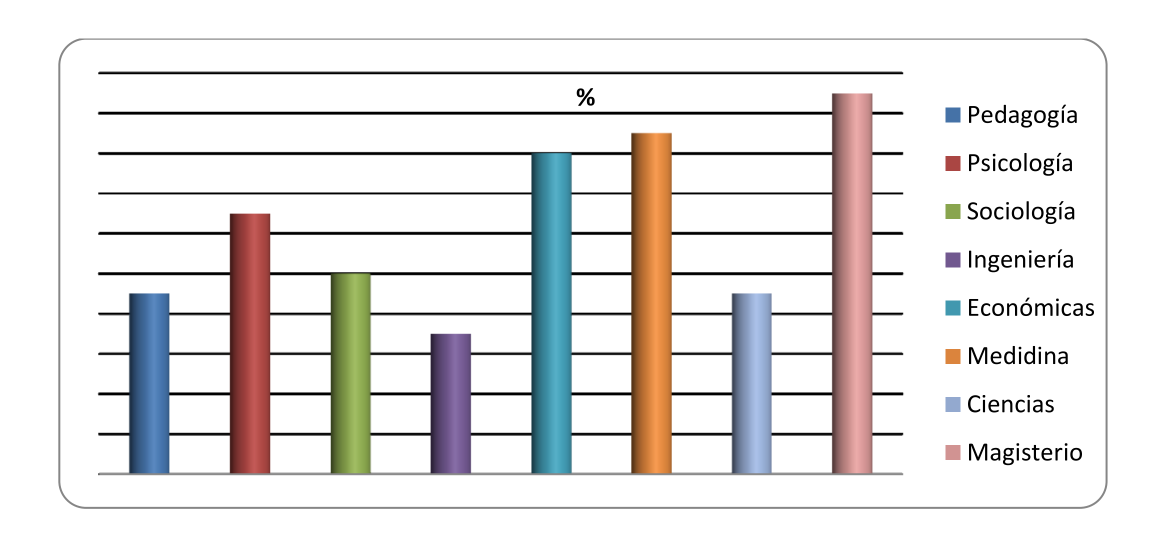


Figura 4c: Diagrama de barras correspondiente a los % de estudiantes según el grado

PARA SU REFLEXIÓN

A la vista de los contenidos de este capítulo, ¿considera que su actitud inicial hacia la asignatura ha mejorado? Si no es así, ¿a qué lo achaca?

¿Considera que el anterior contenido le es de utilidad?

A su juicio, comprender y ser capaz de utilizar los conocimientos del tema, ¿le serán de utilidad cuando comience a estudiar la asignatura?

A su juicio, comprender y ser capaz de utilizar los conocimientos del tema, ¿le serán de utilidad en su vida profesional?

Es probable que haya cuestiones del tema que no haya comprendido. Le invitamos a comunicarlas en el cuestionario de evaluación del curso.

CAPÍTULO 3. REPRESENTACIÓN DE LOS DATOS

Al igual que un alumno puede representar como delegado a su curso o a su carrera, el conjunto de los casos de un grupo, como los 100 anteriores, puede ser representado por uno que pretende ser el que, en conjunto, mejor les representa.

Si una elección de representantes se hace correctamente, es difícil que el representante elegido sea una persona "extrema"; lo habitual es elegir a aquella persona que se aparta menos de la gran mayoría de los casos, razón por la cual representa mejor a todos sin ser igual que ninguno de ellos.

En el manejo de los números que realiza la Estadística* la reducción que hemos presentado –mediante distribuciones de frecuencias, por intervalos más o menos amplios- puede llevarse a manifestaciones superiores cuando todo un conjunto de datos (por ejemplo, los 100 anteriores) se reduce a 1, que pretende representarlos a todos con la mayor fiabilidad.

Cuando decimos que la mayoría de una clase aprueba, que la esperanza de vida de los españoles está en torno a los 80 años, que el abandono escolar supera el 30 %, … estamos representando a todos por ese valor. Es obvio que algunos españoles viven más de 80 años y que la mayoría vive menos, pero ese valor representa mejor que ningún otro a todos los españoles en esa característica.

La Estadística* nos enseña a representar un elevado número de casos por medio de un solo valor para todo un conjunto o grupo. En ocasiones, como veremos, acudimos a tres tipos de valores representativos, que nos ofrecen una visión muy completa del conjunto de puntuaciones del grupo.

Les recomiendo el visionado de la siguiente grabación:

http://www.intecca.uned.es/portalavip/grabacion.php?ID_Grabacion=61218&ID_Sala=65481&hashData=0745b3f95043cc887d78a523a8cabcaf&paramsToCheck=SURfR3JhYmFjaW9uLElEX1NhbGEs

1. Medidas de posición o de tendencia central

Pues bien: cuando en Estadística* se habla de *representación* de un conjunto de datos se piensa generalmente en las medidas denominadas de *posición* o tendencia central*, alguna tan conocida como la *media aritmética**; junto a ella, la *mediana** y la *moda**.

Si en la vida ordinaria se dice de algo que está de moda estamos afirmando que es lo que más se lleva. Por ello, podemos representar los 100 valores anteriores por el que más se da, al que denominamos *Moda* (Mo)* o, como otros dicen, *Modo.* Este valor es el 77, con los datos originales, o el 75,5 (*marca de clase* del intervalo con el mayor número de casos o frecuencia) en la distribución por intervalos.

Otro valor representativo es la *Mediana* (Md).* Basta con ordenar de mayor a menor, o viceversa, la serie original y contar hasta encontrar el que ocupa el lugar central. Si la serie tiene un número par de casos, la Md será el valor medio de los dos centrales. En nuestro caso, con los datos originales, tales puntuaciones son iguales (76) por lo que la Md. coincide con ellos.

Ahora bien: si analizamos la situación, podemos ver que, en el primer caso, solo cuenta la puntuación que más se repite, mientras en el segundo la única que se toma en consideración es la que ocupa el lugar central, sin que ni siquiera importe cuál es su valor. Son dos limitaciones a tener en cuenta.

Ambas limitaciones son superadas por la más completa de estas medidas, le Media o *Media aritmética**, ya que todas y cada una de las puntuaciones de la serie contribuyen a configurarla en proporción a su valor. Por ello, para su cálculo no importa cuál sea la más repetida o cuál ocupe un determinado lugar en la serie ordenada; de hecho, no es preciso ordenar la serie sino sumar todas las puntuaciones y dividir la suma por el número de casos (N). Para el cálculo de la Media se aplica la ecuación siguiente:

$$Ecuación\ 3: Media = \frac{\sum_{i=1}^{N} X_i}{N}$$

El símbolo $\sum_{i=1}^{n} X_i$ debe leerse como sigue: súmense todas las puntuaciones X desde la puntuación i a la puntuación N, esto es, desde la primera a la última.

En el supuesto de calcular la media en una distribución de frecuencias, la anterior ecuación se convierte en esta otra (ecuación 4), donde el valor X_i no es una puntuación directa sino la marca de clase del intervalo:

$$Ecuación\ 4: Media = \frac{\sum_{i=1}^{N} x_i f_i}{N}$$

Compruebe el lector la pequeña distorsión que se da entre este valor, 73,02, el más exacto, y el obtenido en el caso de la distribución de 11 intervalos, donde la suma de los productos de las marcas de clase por sus frecuencias arroja un valor muy próximo: 7230, con la cual la media es de 72,3. Puede comprobar estos datos en la tercera fila de la tabla 3 y en la última columna.

El tipo de medidas que se utiliza más comúnmente para representar a un grupo es el de tendencia central o posición y, dentro de estas, la media aritmética* es la más completa; pero solo debe utilizarse con variables medidas con escalas de razón o cociente y de intervalo. En ocasiones, cuando los rangos de una variable ordina*l se aproximan razonablemente a una escala de intervalo, también se suele utilizar la media aritmética*.

2. Medidas de dispersión* o variabilidad

Ponga ahora atención el lector a estas dos series de datos ya presentados anteriormente:

5, 5, 5, 5, 5, 5, 5, 5, 5, 5
10,10, 10, 10, 10, 0, 0, 0, 0, 0

Si calculamos la Mediana*, en ambos casos es la misma: 5 en la primera serie y (10 + 0) : 2 = 5 en la segunda. Y si lo hacemos con la Media, en ambos casos obtenemos una media de 5.

Sin embargo, a nadie se le oculta que estamos ante dos conjuntos de datos radicalmente diferentes, a pesar de que el valor representativo Media sea el mismo. Para hacer más realista el caso, piense en un profesor que tiene no 10 alumnos sino 20 o 30, en dos clases distintas: en la

primera, los 20 o 30 niños, con puntuaciones de 5 en Matemáticas y en la segunda, con la mitad de casos con 10 y la otra mitad con 0. Parece obvio que no debería actuar del mismo modo en ambas clases.

Un tipo de medidas representativas diferente del anterior (medidas de posición o tendencia central) es el denominado de ***dispersión****, que nos informa de esta característica. Si en la primera de las dos series anteriores la dispersión en nula, dado que todas las puntuaciones coinciden con la Media, en el segundo es máxima ya que todos los casos se sitúan en los extremos.

En un caso como este, basta fijarnos en lo que se conoce como *rango* de la serie para hacernos una idea clara del grado de dispersión*. Pero lo representado en ambas series no es lo habitual. Ni, por lo general, todos obtienen la misma puntuación ni se da una fractura tan grande entre los miembros del grupo.

Para apreciar la magnitud de la dispersión* contamos con medidas específicas, tales como la *desviación mediana, la desviación media, la desviación típica* o la varianza**.

El mismo nombre de la primera –*desviación mediana*- ya nos sugiere en qué consiste: es la media de las desviaciones de las puntuaciones con respecto a la Md del grupo. En el caso de la desviación media se trata, también, de la media de las desviaciones, pero ahora tomando como referencia la media aritmética*.

Ahora bien: podemos comprobar qué es lo que pasa cuando hacemos estas operaciones en la siguiente serie, donde la media 5: (50 : 10) y Md es 6 (Tabla 4)

X_i	1	1	2	3	6	6	7	7	8	9	$\sum$= 50
X_i – Md	-5	-5	-4	-3	0	0	1	1	2	3	-10
$\|X_i$-Md$\|$	4	4	3	2	1	1	2	2	3	4	26
X_i-Media	-4	-4	-3	-2	1	1	2	2	3	4	0
$(X_i\text{-Media})^2$	16	16	9	4	1	1	4	4	9	16	80

Tabla 4: Tratamiento de los datos (X_i) para el cálculo de medidas de dispersión

Como se puede apreciar, en el primer caso obtenemos una suma positiva o negativa según que la distribución tienda a los valores inferiores o superiores a la Md (en este caso, los valores son negativos). Pero en el segundo la suma da, y siempre dará, 0 como consecuencia de las propiedades de esa medida de posición. Por eso, en el caso de la desviación mediana tendremos que tomar las desviaciones en valor absoluto (lo que se representa por el símbolo | |) y trabajar con la suma de las mismas

Ecuación 5: $DMd = \frac{|\sum_{i=1}^{N} X_i - Md|}{N}$

DMd = 26 / 10 = 2,6

No obstante, no es esta la medida de dispersión* más utilizada. Siempre que es posible, se acude a la desviación típica*, representada por *s*, y a su cuadrado, conocido como varianza*, representada por s^2.

En ambos casos, las desviaciones con respecto a la Media (X_i – Media) se elevan al cuadrado a fin de evitar que la suma dé 0. Pues bien: la *varianza* (s^2) es la media de las desviaciones de las puntuaciones individuales con respecto a la media, elevadas al cuadrado; por su parte, la desviación típica* (s) es la raíz cuadrada de la anterior.

Ecuación 6: $s = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Media)^2}{N}} = \sqrt{\frac{80}{10}} = 2{,}828$

Ecuación 7: $s^2 : \frac{\sum_{i=1}^{N}(X_i - Media)^2}{N} = 8$

Junto a las medidas de posición, podemos caracterizar un grupo con las de ***dispersión* o variabilidad***, que nos ofrecen una idea del grado de concentración de las puntuaciones directas en torno a la media, lo que tiene evidentes aplicaciones para la práctica profesional. Hemos citado, como fundamentales, la desviación media, la desviación típica* y la varianza*.

Estas medidas tienen su uso más frecuente en la denominada Estadística inferencial*; una utilidad muy común e importante es la de interpretar una puntuación individual en el marco de una distribución normal (campana de Gauss) como veremos más adelante.

Suponiendo que nuestra distribución empírica de datos se acomoda al modelo normal podremos interpretar la puntuación de un sujeto cualquiera viendo cuántas unidades de s se aparta de la media del grupo, algo que podemos traducir fácilmente a porcentajes como tendremos ocasión de ver.

Esa puntuación individual*, basada en s, se conoce como puntuación típica* (z) a la que ya nos hemos referido, e indica en cuántas desviaciones típicas se aparta un sujeto de la media del grupo (Ecuación 2).

Aunque tendremos ocasión de verlo con más detalle, lo podemos apreciar en el siguiente gráfico de la curva normal de probabilidades* (Figura 5):

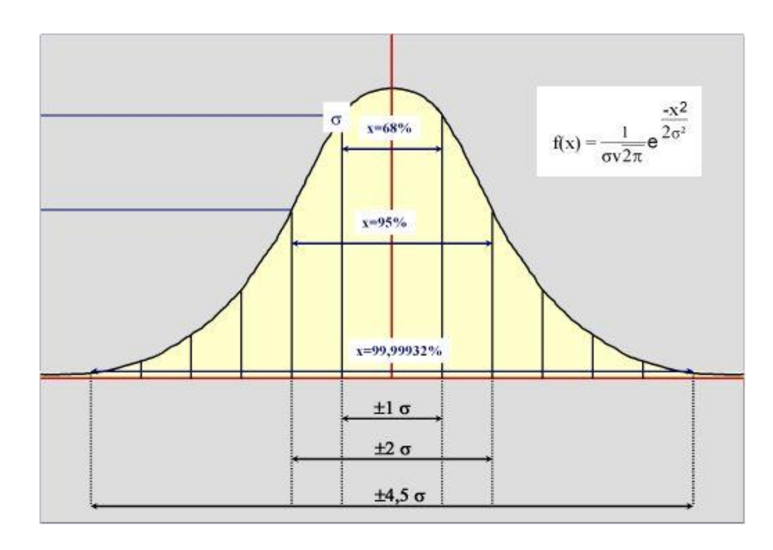


Figura 5: Curva normal de probabilidades o Campana de Gauss

Cualquier puntuación individual (X_i)* ocupa un lugar en la curva, por encima o por debajo de la ordenada de la Media (línea roja vertical), que la divide en dos partes simétricas. Las puntuaciones cercanas a la Media se encuentran a su derecha o a su izquierda, según sean mayores o menores que ella. Una puntuación X_i que se aparte una desviación típica* por encima o por debajo de la media se situará en la ordenada correspondiente del gráfico (± σ). Pero de esto hablaremos más adelante.

Baste decir ahora que la Estadística* hace sus verdaderas aportaciones en lo que denominamos *inferencia*, que no es sino el proceso por el cual *estimamos* determinados valores de una variable en el conjunto total de casos (*población**) a partir de los *medidos* en una *muestra** de la misma. Los valores *medidos* en la muestra se denominan estadísticos* y se representan como hemos hecho hasta ahora (M, Md, Mo, DMd, s, s^2, …) Los valores *estimados* en la población se denominan parámetros* y para ellos utilizamos letras griegas (para el parámetro Media utilizamos μ, para la desviación típica, σ)

Un ejemplo claro y sencillo: un profesor con 4500 puede tomar una muestra* de los mismos de 150, obtener su media y estimar cuál será la media (μ) de los 4500. Y lo mismo con la desviación típica (σ)

Otro: en las encuestas sobre intención de voto, se suelen tomar muestras de no más de 2 o 3 mil sujetos; a partir de sus respuestas se estima la intención de voto de los varios millones de españoles que votarán.

Sin entrar en detalles, se comprende:

a) *Que los datos fiables son los **medidos** en la muestra**
b) *Que los datos **estimados** en la población* podrán apartarse en mayor o menor grado del verdadero valor.*
c) *Que la precisión de la estimación depende de la calidad de la muestra**
d) *Que los datos más útiles son los **estimados*** a pesar del error de estimación que les afecte.

Cuanto más seguridad desee el investigador para sus estimaciones, más calidad deberá tener su muestra*, esto es: más *representativa* de la población*, lo que exige un *tamaño* suficiente y una selección *imparcial* de los sujetos, por lo general *aleatoria.* Para hacernos una idea de lo que entendemos por representatividad podemos acudir a una fotografía con respecto a la persona. Las fotografías pueden ser más o menos fieles al sujeto fotografíado.

Pues bien: para esos procesos de inferencia, las medidas de dispersión* más utilizadas son la *varianza** y la *desviación típica*.* Su cálculo es sencillo a partir de los datos de la tabla 4, ya que no es sino la media de las desviaciones elevadas al cuadrado, en el primer caso; en el segundo, es la raíz cuadrada de dicho valor. Cuestión diferente, como veremos, es la de su interpretación.

En nuestro caso, tal suma alcanza el valor de 80, por lo que la varianza* será:

$$s^2 = \frac{(\sum(XI - Media)^2}{N} = 80 : 10 = 8,$$

y la desviación típica $s = \sqrt{\frac{80}{10}} = 2{,}828$

Preciso es reconocer que no resultan de fácil comprensión ambos conceptos. Asumamos la idea de que se trata de la media de las desviaciones con respecto a la media (en el caso de la varianza*), y de la raíz cuadrada de esta en el segundo.

Pero avancemos la importancia que tendrá la segunda cuando iniciemos el estudio de los modelos de probabilidad*, como es la curva normal* o campana de Gauss, de gran importancia y uso (la desviación típica) o las pruebas de significación estadística, como la prueba F, para decidir si es razonable o no tomar en consideración determinadas diferencias (la varianza).

Un problema de estas medidas es su difícil interpretación; ni es fácil decidir sobre el grado de dispersión* de una serie (si es poca, media o elevada) salvo si fuera nula, cuyo valor es 0, ni, mucho menos, decidir si una serie es más o menos dispersa que otra. A este último aspecto daremos respuesta mediante el *coeficiente de variación.*

Por el momento, dejémoslo ahí y avancemos con otras medidas de dispersión*, como el *recorrido semiintercuartílico y el coeficiente de variación.*

Si la desviación típica* –s- se utiliza mucho en la estadística descriptiva, la varianza* –s^2- ofrece grandes aplicaciones en la inferencial.
Otras medidas a tener en cuenta son el recorrido intercuartílico –el que va entre los cuartiles 1 y 3- y el semiintercuartítico.

Ya conocemos la Mediana*, medida de posición. Pues sepamos que la Md, que deja por encima y por debajo de sí al 50 % de los casos, equivale a lo que denominamos cuartil 2 (Q_2 = Md). Si cada una de las mitades se divide a su vez en partes iguales, la serie total queda dividida en cuatro partes mediante tres cuartiles: Q_1, Q_2, Q_3. Pues bien, el 50 % central de la serie se denomina *recorrido intercuartílico*, y su división por 2 *recorrido semi-intercuartílico.*

Su valor nos da información sobre la dispersión de la serie, como fácilmente se desprende de las tres siguientes series de datos: no es lo mismo que en una serie el 50 % central se encuentre ente puntuaciones muy próximas que el que para reunir ese 50 % tengamos que apartarnos ampliamente de la mediana del grupo. Veamos las tablas 5, 6 y 7:

X_i	1	2	4	5	6	7	8	10	N
f_i	2	3	4	7	6	5	2	1	30
f_a	2	5	9	16	22	27	29	30	

Tabla 5: Distribución de frecuencias (f_i) y de frecuencias acumuladas (f_a)

Sin entrar en detalles, la Md es 5; y los Q_1 y Q_3 4 y 7. Por tanto, el 50% de los casos se encuentra entre 4 y 7, siendo ese el valor de tal recorrido. Lo podemos apreciar fácilmente si la serie anterior la convertimos en datos originales, sin agrupar por frecuencias:

1,1,2,2,2,4,4,4,4,5,5,5,5,5,5,5, 5,6,6,6,6,6,6,7,7,7,7,7,8,8,10

Q_1 Md = Q_2 Q_3

X_i	1	2	4	5	6	7	8	10	N
f_i	3	5	5	4	5	3	2	3	30
f_a	3	8	13	17	22	25	27	30	

Tabla 6: Distribución de frecuencias (f_i) y de frecuencias acumuladas (f_a)

Aquí Q_1 y Q_3 son 2 y 6, respectivamente; por tanto, la serie presenta mayor dispersión*; es más plana que la anterior, que tiene mayor apuntamiento en los valores centrales. Veámoslo con datos sin agrupar:

1,1,1,2,2,2,2,2,4,4,4,4,4,5,5, ,5,5,6,6,6,6,6,7,7,7,8,8,10,10,10

Q_1 Md = Q_2 Q_3

X_i	1	2	4	5	6	7	8	10	N
f_i	8	6	3	2	2	1	6	2	30
f_a	8	14	17	19	21	22	28	30	

Tabla 7: Distribución de frecuencias (f_i) y de frecuencias acumuladas (f_a)

Aquí Q_1 y Q_3 son 1 y 8, respectivamente; por tanto, la serie presenta todavía mayor dispersión* que la anterior, siendo más plana que las dos anteriores, que tienen mayor apuntamiento en los valores centrales.

1,1,1,1,1,1,1 1,2,2,2,2,2,2,4, 4,4,5,5,6,6,7,8,8,8,8,8,8,10,10

Q_1 Md = Q_2 Q_3

Podemos "ver" de modo más intuitivo, mediante representaciones gráficas lo que representa esta medida de dispersión* en gráficos como el diagrama de caja (figura 6a):

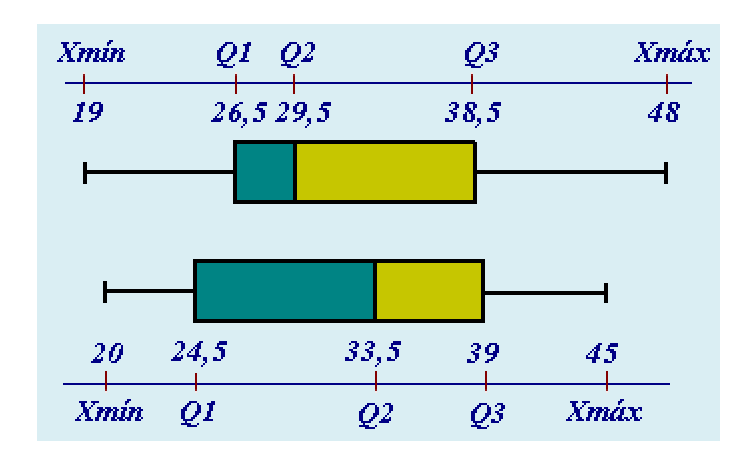


Figura 6.a: Diagrama de caja y bigotes

La importancia de ese tipo de gráfico es la gran información que contiene:

- *Las puntuaciones extremas (19 y 48;20 y 45*
- *La Mediana, igual al Q_2 o cuartil 2 (29,5 y 33,5)*
- *El recorrido intercuartílico: Q_1 a Q3: 38,5 – 26,5; 39 – 24,5.*

- *La información contenida sobre este valor en el rectángulo central: en él podemos apreciar si se da equilibrio o no entre los diferentes cuartiles; si no se da, como ocurre en este caso, si los valores predominantes están por encima o por debajo de la mediana o Q_2; del mismo modo, podemos apreciar lo que ocupa el 50 % central y lo propio de las puntuaciones extremas (los "bigotes") superior en el primero de los casos.*

En la figura 6.b podemos apreciar la facilidad que permite esta representación para comparar diferentes aspectos de las respuestas de una muestra de profesores en un cuestionario sobre la valoración que realizan en torno a la organización metodología, evaluación entre otros.

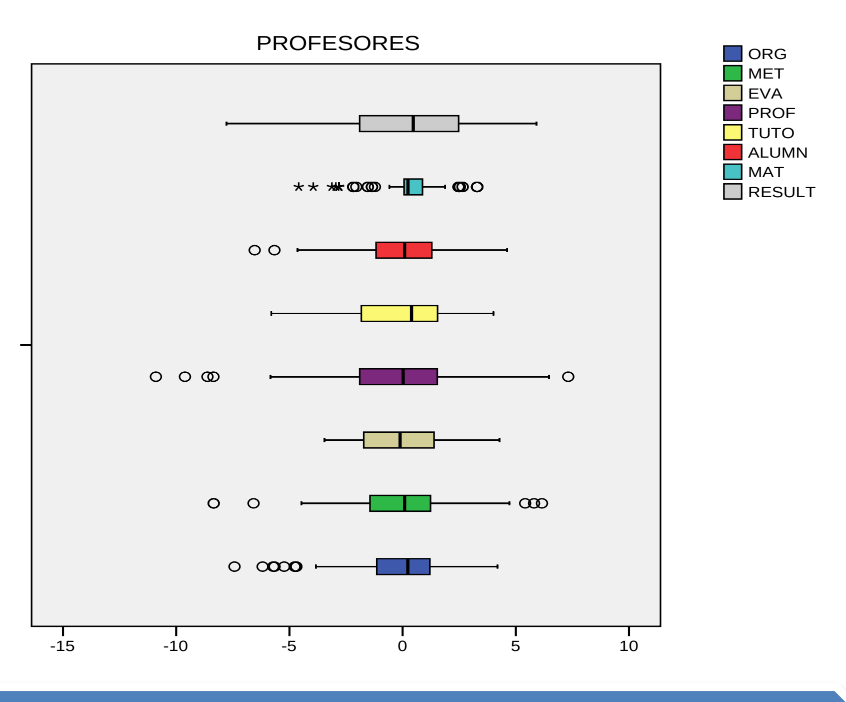


Figura 6.b: Diferentes diagramas de caja y bigotes

Terminaremos este punto con el *coeficiente de variación*, una sencilla medida que no es sino el cociente entre la desviación típica* y la media del grupo. Su principal utilidad es la de facilitar la comparación de la dispersión de dos series de datos.

$C.V. = s/\text{Media}$; en nuestro caso estaríamos ante 5 : 8,94 = 0,56

La interpretación de las medidas de dispersión* es más difícil que la correspondiente a las de posición. Lo que puede ayudarnos, cuando la distribución es compatible con la normal, es el uso de este modelo.
Existe una medida que nos facilita la comparación de la dispersión* o variabilidad de varias series de datos; se trata del **coeficiente de variación**.

3. Medidas de forma

En nuestro recorrido por las medidas de representación hemos visto las de *posición* o de *tendencia central* y las de *dispersión**.

Utilizadas conjuntamente, tenemos una valiosa información para hacernos una idea de las características de un grupo. Pero podemos mejorar tal información mediante otras dos medidas de interés, no tanto por sus propias aportaciones como por lo que contribuyen a la caracterización del grupo; nos referimos a las de *simetría* y de *apuntamiento,* denominadas en algunos manuales como medidas de *forma* por ofrecer información sobre la forma general de la distribución de los datos.

Veamos estas series de datos (Series 3 a, b, c, d, e):

a) 5, 5, 5, 5, 5, 5, 5, 5, 5, 5
b) 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
c) 9, 9, 9, 9, 9, 9, 9, 9, 9, 9
d) 1, 2, 4, 4, 5, 5, 6, 6, 8, 9
e) 1, 1, 1, 3 5, 7, 8, 9, 9, 9

Las tres primeras series tienen la misma forma, una forma uniforme o plana; la diferencia se da en que los valores son medios en a) y extremos en b) y en c). La serie d) es más habitual: los valores extremos son menos frecuentes que los medios. Y la serie e) presenta una distribución menos frecuente, con más casos en los extremos que en el centro.

Si centramos nuestra atención en d) observamos que el valor más frecuente, el 5, está en el centro, y que tiene tantos valores a su izquierda como a su derecha; además, sus frecuencias descienden hacia ambos extremos en la misma forma: 2, 1 y 1 casos. Si representáramos la serie y la dobláramos por la mitad apreciaríamos su *simetría.*

Las medidas de forma nos ofrecen una idea de dos características del grupo como tal: el grado en que se acercan a la ***simetría***, característica del modelo normal, y el de ***apuntamiento***, más o menos equilibrado.

3.1. Simetría / asimetría

Pues bien; una medida de *forma* es la que nos indica su simetría o, mejor, el grado de *asimetría* de una distribución empírica; se representa por g_1 y mide el grado de *asimetría* de una serie de puntuaciones, esto es: la medida es que la serie empírica se aparta de una distribución *simétrica,* característica propia de las distribuciones denominadas *normales*, esto es, de las que siguen el modelo de la denominada *curva normal de probabilidades* o campana de Gauss*, una de cuyas características definitorias es la de ser simétrica con relación a la ordenada de la media.

La medida del grado de asimetría, denomina *coeficiente de asimetría*, se representa por g_1 y se obtiene mediante la siguiente ecuación:

Ecuación 8: $$g_1 = \frac{\sum_{i=1}^{N}(xi - Media)^3}{Ns^3}$$

Cuando el número de valores de una distribución es mayor en la parte inferior a la media que en la superior a la misma, la distribución se muestra *asimétrica* hacia la izquierda, y hacia la derecha en caso contrario. En el primer caso $g_1 < 0$ y la asimetría se considera negativa; en el segundo, $g_1 > 0$, y la asimetría es positiva.

Si las diferencias entre los valores positivos y negativos en (X_i – media) tienden a 0, la distribución se considera simétrica. La elevación de este valor al cubo se debe a que se trata de evitar que la $\sum(Xi - media) = 0$, como nos ocurría en el caso de la varianza*. En la figura 6 se presentan sendos ejemplos:

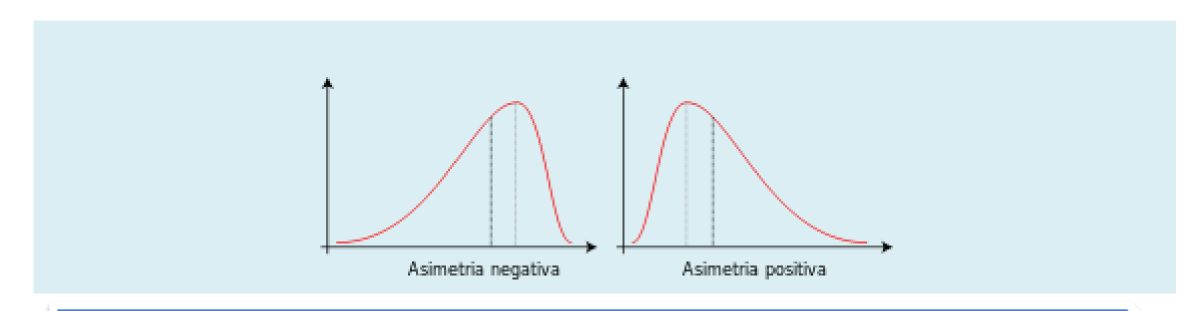


Figura 6: Distribuciones con asimetría positiva (hacia la derecha) y negativa (hacia la izquierda)

Las medidas de asimetría nos permiten calificar la distribución de las puntuaciones de un grupo como "normal" o como asimétricas, en mayor o menor grado, bien sea asimetría positiva o negativa.

3.2. Apuntamiento o curtosis

También con la serie d) podemos cuantificar su *apuntamiento* (simbolizado por g_2) esto es: el grado en que las puntuaciones centrales se concentran en torno a la media del grupo. El *apuntamiento* también recibe el nombre de *curtosis.*

Sin entrar en explicaciones que no vienen al caso, diremos que el apuntamiento *normal* se representa por $g_2 = 3$; valores de $g_2 > 3$ representan una distribución que recibe el nombre de *leptocúrtica*, mientras que en el caso de distribuciones con $g_2 < 3$, más achatadas, la distribución se denomina *platicúrtica.* La normal, obviamente, recibe el nombre de *mesocúrtica.*

La distribución *leptocúrtica* no solo tiene un mayor apuntamiento central sino que los valores extremos presentan, también, mayores frecuencias que en la normal. Por tanto, si un profesor está ante una distribución leptocúrtica sabe que sus alumnos se concentran más en el centro que en los extremos y que las puntuaciones extremas presentan frecuencias más elevadas que las que se darían si la distribución fuera normal.

El apuntamiento se obtiene mediante:

$$Ecuación\ 9\text{: } g_2 = \frac{\sum_{i=1}^{N}(X_i - Media)^4}{ns^4}$$

En la figura 7 pueden apreciar curvas con diferente grado de apuntamiento, superior e inferior al normal.

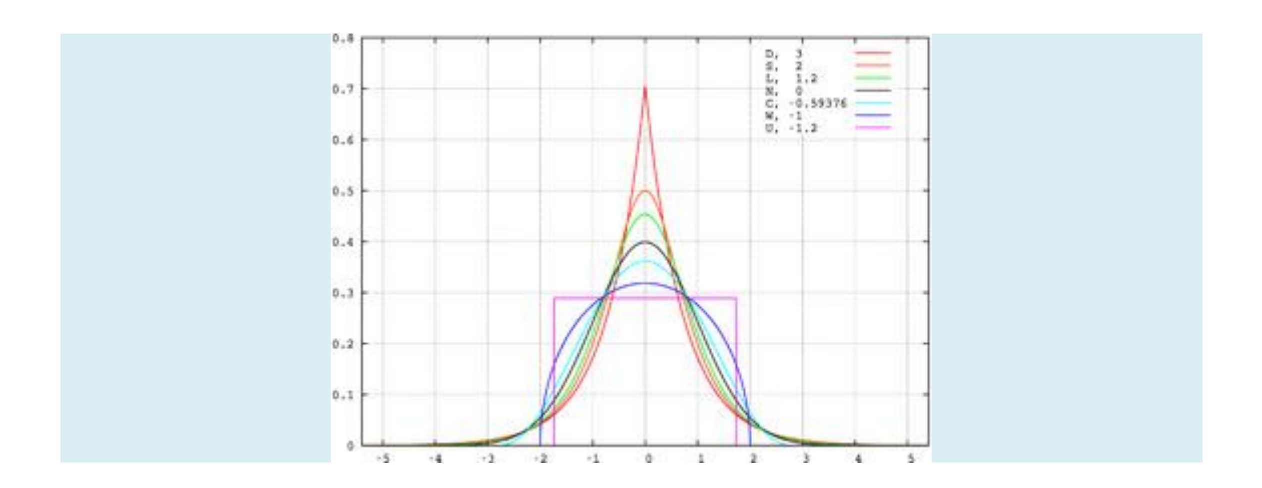


Figura 7 Distribuciones leptocúrticas, mesocúrtica y platicúrticas

Las medidas de apuntamiento nos permiten calificar la distribución de las puntuaciones de un grupo como mesocúticas o normales o bien como leptocúrticas o platicúrticas en diverso grado.

PARA SU REFLEXIÓN

A la vista de los contenidos de este capítulo, ¿considera que su actitud inicial hacia la asignatura ha mejorado? Si no es así, ¿a qué lo achaca?

¿Considera que el anterior contenido le es de utilidad?

A su juicio, comprender y ser capaz de utilizar los conocimientos del tema, ¿le serán de utilidad cuando comience a estudiar la asignatura?

A su juicio, comprender y ser capaz de utilizar los conocimientos del tema, ¿le serán de utilidad en su vida profesional?

Es probable que haya cuestiones del tema que no haya comprendido. Le invitamos a comunicarlas en el cuestionario de evaluación del curso.

SINTESIS

Con el ánimo de recapitular los contenidos presentados hasta aquí, se ofrece un resumen de los aspectos fundamentales.

La Estadística y los números

- La Estadística* es una ciencia que trabaja con números.
- Los números se obtienen a partir de medir, pesar o contar objetos, sean estos directamente observables o no; por objeto entiéndase cualquier realidad que pueda medirse, pesarse o contarse, tanto si es animada como inanimada, si es persona o cosa, si es directamente accesible como si no.
- La calidad de los números depende, fundamentalmente, de la posibilidad de aplicar a los objetos medidos unidades de medida fiables y válidas.
- Frente a ciertas medidas referidas a objetos directamente accesibles, como la edad, el peso, la talla…, en nuestros ámbitos debemos acudir a "objetos" no directamente observables, lo que exige su definición (una definición denominada por lo general *constructo*) y la construcción de instrumentos adecuados para atribuirles valores. Para construirlos se procede a la denominada definición *operativa* de la variable.
- Según sean los números obtenidos de las medidas de los objetos será o no lícito aplicarles ciertas propiedades y utilizarlos en determinadas operaciones matemáticas. Los más perfectos pertenecen a las denominadas escalas de *razón o cociente*, seguidos por los de escala de *intervalo*, de las *ordinales* y, por último, de las *nominales.*

Interpretación de los números

- Los números obtenidos a partir de la utilización de instrumentos esconden información que es preciso extraer.
- Los números referidos a sujetos concretos (alumnos, pacientes, partidos políticos, desempleados…) no son fácilmente interpretables en sí mismos. Podemos hacerlo si conocemos el "suelo" y el "techo" (puntuaciones mínima y máxima) del instrumento de recogida de datos y la unidad de medida, pero es más frecuente situar esas puntuaciones en el conjunto del grupo del que forma parte.
- Entre las medidas individuales más habituales, podemos citar la de desviación (X_i – *Media aritmética* = x_i)*, la puntuación típica (z_i)*, los cuantiles (cuartiles, deciles, percentiles), la Edad Mental (EM) o el cociente intelectual (EM / Edad cronológica).
- Entre los procedimientos más sencillos para extraer información de los números se encuentra la simple *ordenación* de los mismos.
- Cuando el conjunto de valores de un grupo es elevado, la simple ordenación puede no ser suficiente para apreciar las características que lo definen.
- Entonces podemos acudir a su *reducción*, haciendo que, sin alterar, o alterando de modo mínimo los datos originales, podamos hacernos una idea de las características del grupo con unos pocos valores.
- Esta forma de actuar consiste en construir *distribuciones de frecuencias* en las que cada valor (X_i) va acompañado del número de veces que aparece (frecuencia: f_i).

- Cuando estas distribuciones mantienen todo los valores originales, la distribución no se altera en absoluto. Sin embargo, en ocasiones, cuando la serie tiene un muy amplio *recorrido* (distancia entre los valores máximo y mínimo) puede ser conveniente que la distribución se reduzca construyendo *intervalos* de amplitud mayor que 1 (I_i) incluyendo para cada intervalo el número de casos –frecuencia- del conjunto de puntuaciones del intervalo.
- En estos casos, la denominada *marca de clase* o valor medio del intervalo (X_i) se toma como representativa del intervalo a los efectos de los cálculos, lo que puede representar pequeñas desviaciones –positivas o negativas- entre los resultados de los cálculos con datos originales o de una distribución de esta naturaleza.
- Estas desviaciones, por lo general, serán pequeñas porque, habitualmente, las diferencias positivas en unos casos se compensarán con las negativas en otros.
- Teniendo en cuenta las limitaciones de los datos en fiabilidad y validez estas pequeñas desviaciones no deberían preocuparnos. La apariencia de de exactitud que nos da una calculadora con muchos decimales no refleja la realidad de los valores medidos, afectados por las limitaciones de los instrumentos de medida.

Caracterización y representación de grupos

- La reducción de datos puede suponer una notable simplificación de los datos originales, haciéndolos más manejables; pero la Estadística nos permite algo más: *representar* el conjunto por medio de unas medidas que nos informan de las características más importantes del conjunto de datos.
- Como toda *representación* nunca será tan perfecta con los datos originales, pero mientras estos, si son numerosos, se hacen muy difíciles de comprender y de tratar, aquellos los representan con la calidad suficiente para comprender la naturaleza y características del conjunto.
- Tres son los tipos de medidas que nos ayudan a comprender las características de un grupo (Cuadro 5):

POSICIÓN O TENDENCIA CENTRAL	DISPERSIÓN*	FORMA	
		SIMETRIA	APUNTAMIENTO O CURTOSIS
Moda: Mo	Recorrido	g_1	g_2
Mediana: Md	Desviación media	$g_1 > 0$. Asimetría positiva	$g_2 > 3$. Leptocúrtica
Media: M	Desviación típica: s	$g_1 = 0$. Simétrica Normal	$g_2 = 3$. Mesocúrtica
	Varianza: s^2	$g_1 < 0$. Asimetría negativa	$g_2 < 3$. Platicúrtica
	Recorrido semi-intercuartílico		
	Coeficiente de variación		

Cuadro 5 Medidas representativas de grupo

- Las medidas de *posición* nos informan sobre la tendencia de la distribución de datos a acumularse en el centro de la misma (de ahí su otra denominación: de tendencia central)

    - Entre las medidas de posición, la más perfecta es la Media aritmética* (por lo general denominada Media), dado que en ella influyen, de modo proporcional a su valor, todas y cada una de las puntuaciones de los datos originales. Resulta especialmente adecuada para medidas de razón o de intervalo.
    - Le mediana* (Me o Md según los textos) también es una importante medida, pero tiene como inconveniente que en ella las puntuaciones no influyen por su valor sino por el lugar que ocupan, de modo que series muy diferentes pueden tener la misma mediana con solo mantener la misma puntuación central. Está especialmente adecuada a medidas de escala ordinal.
    - La Moda* o Modo (Mo), poco utilizada, solo indica el valor más repetido. Se aplica fundamentalmente a puntuaciones de escala nominal.

- Las medidas de *dispersión** son, probablemente, las más relevantes en el análisis de los datos numéricos, especialmente en la Estadística inferencial. Nos informan sobre el grado en que las puntuaciones se concentran o se separan de la media del grupo. En Estadística la *dispersión* de las puntuaciones es una cualidad o característica de gran valor y utilidad, como tendremos ocasión de ver.

    - Las más importantes son las más abstractas, en concreto la *varianza** (s^2) o media de las desviaciones de las puntuaciones con respecto a la media, elevadas al cuadrado, y la *desviación típica** o raíz cuadrada de la anterior.
    - En sí mismas nos ofrecen una información valiosa sobre la concentración o dispersión de las puntuaciones de una serie, si bien su interpretación no es fácil.
    - Además, estas dos medidas se utilizan mucho en la *inferencia estadística*, proceso por el cual *estimamos* los valores que se darán en la *población** (conjunto total de datos) a partir de los *medidos* en una *muestra** de la misma, Los valores *medidos* se denominan *estadísticos** (media, desviación típica, varianza…) y los *estimados* se denominan *parámetros** (estos se representan mediante las correspondientes letras griegas: μ, σ, σ^2…).
    - Para *estimar* los parámetros* tendremos que servirnos de los *modelos estadísticos** y de la teoría de la *probabilidad**. De este modo, cualquier valor estimado vendrá acompañado de la probabilidad* de que ocurra.
    - El modelo de referencia más habitual es el denominado *normal*. Tomándolo como referencia, decidimos si la distribución empírica es platicúrtica o leptocúrtica, si su asimetría es negativa o positiva.
    - Contamos con pruebas que nos permiten decidir si una distribución empírica se acomoda o no a la normal; en caso positivo, podemos aplicar a los datos empíricos las propiedades del modelo, pensando que las desviaciones apreciadas se deben a pequeñas imperfecciones en la selección de los datos.
    - En este proceder no hacemos sino algo habitual: nadie ha visto en la Naturaleza un cono, pero sí montañas más o menos cónicas (pensemos en el Teide). Pues bien: dando por bueno que el Teide no se aparta mucho de un cono ideal, podemos calcular, aproximadamente, su superficie y su volumen, aplicándole la fórmula del modelo, del cono.
    - Como se puede comprender, el problema es decidir si el objeto empírico se acomoda *razonablemente* al modelo; la Estadística nos ayudará a ello mediante

pruebas denominadas de *bondad de ajuste** (por ejemplo, para el caso del ajuste a la curva normal, la de χ^2; léase ji o chi cuadrado)

- Las medidas de *forma*, como su nombre indica, nos ofrecen una visión global sobre la forma de la distribución, fijándose en dos aspectos fundamentales: la *simetría* y el *apuntamiento.* Para valorar tales características se toma como referencia la denominada *distribución normal*, que es *simétrica* respecto de la ordenada de la media y que tiene un apuntamiento normal –*mesocúrtica*- en sus valores centrales.
    - La *asimetría* puede ser negativa, cuando el valor del correspondiente coeficiente tiene valores negativos, quedando sesgada hacia la izquierda, o positiva, cuando el correspondiente valor es positivo, quedando sesgada hacia la derecha.
    - El apuntamiento normal nos sitúa ante distribuciones *mesocúrticas*, siendo *leptocúrticas* cuando el apuntamiento es mayor y *platicúrticas* si es menor.

- Si se ofrecen datos de estos tres tipos de medidas, la caracterización de una distribución de puntuaciones es muy completa y, sobre permitirnos una comprensión profunda de sus características, nos facilitará la realización de determinados procesos de inferencia, entre los que destacamos, precisamente, la *estimación de parámetros**, con determinada probabilidad*, y la realización de *contrastes*, mediante pruebas estadísticas que nos permitirán tomar decisiones sobre los efectos de las variables independientes* sobre las dependientes* en los experimentos*.

- Por otra parte, los seres humanos estamos más habituados a comprender los fenómenos que ocurren ante nuestros ojos o que somos capaces de representar de forma intuitiva.

- Pues bien: los números también pueden representarse mediante una serie de representaciones, que vamos a ver, y que nos facilitan la interpretación de forma más fácil; digamos, no obstante, que los números, unidos a sus representaciones gráficas, se complementan: estas ofrecen la visión intuitiva; aquello, la precisión.

PARA SU REFLEXIÓN

¿Considera que el anterior resumen de lo estudiado le es de utilidad?

Es probable que, a pesar del repaso, haya cuestiones del tema que no haya comprendido. Puede manifestarlo en el cuestionario de evaluación del curso.

CAPÍTULO 4. EL CASO DE DOS O MÁS VARIABLES

La información contenida en una serie de datos puede resultar de sumo interés para *comprender* la naturaleza y características del grupo al que hace referencia.

Sin embargo, es preciso reconocer que los fenómenos humanos son muy complejos, por lo que es frecuente que entre los intereses de los profesionales o de los estudiosos se encuentre el de conocer la *relación** o falta de ella (independencia) entre dos o más series de datos.

Preguntas tales como: ¿está relacionada la inteligencia con el sexo, la raza o la escolaridad?. ¿Mantienen relación las técnicas de estudio con las calificaciones? ¿Qué relación se da entre el número de horas de estudio y los resultados académicos? ¿Se relaciona la violencia juvenil con el analfabetismo? ¿Hay relación entre el autoconcepto y la asertividad?. ¿Se da relación entre el consumo de estupefacientes y el nivel cultural?...

A priori, cabe pensar que a más horas de estudio, mejores resultados, pero ¿no puede ocurrir que, a partir de cierto número de horas el aprendizaje baje y hasta sea nulo? ¿No puede ser que la relación varíe según el tipo de aprendizaje, memorístico o comprensivo? ¿O que dependa del momento del día: por la mañana, a medio día o por la tarde?

En cuanto a la relación inteligencia - sexo ¿podría variar según el tipo de inteligencia de que se trate (recordemos a Gardner y sus inteligencias múltiples)? ¿Podría ocurrir lo mismo con la raza?

> Conocer si dos o más variables* ***co-varían***, esto es, varían conjuntamente en una u otra dirección, es una información valiosa para el ejercicio profesional en ámbitos diversos. En tales casos se dice que las variables están relacionadas y su relación se denomina en Estadística ***correlación****.

Les recomiendo visualizar la siguiente grabación:

http://www.intecca.uned.es/portalavip/grabacion.php?ID_Grabacion=61401&ID_Sala=65580&hashData=6f73b314d5f8b36b108763b24dc6a805

1. La correlación*. Tipos y valores

Pues bien: cuando disponemos de dos series de datos y deseamos responder a preguntas de ese tipo, la Estadística acude en nuestro auxilio al permitirnos establecer si se da o no relación, denominada aquí *correlación** y representada de ordinario por r_{XY} (se lee correlación entre las variables X e Y) de qué tipo (positiva o negativa) y con qué intensidad (perfecta o imperfecta).

La existencia de relación supone que las dos series de datos co-varían, esto es: varían conjuntamente; si hay correlación, el hecho de que los valores de una aumenten o disminuyan implica que los de la otra aumentan o disminuyen (correlación positiva) o bien que disminuyen o aumentan (correlación negativa).

Si los cambios mantienen una misma proporcionalidad, la correlación será perfecta, y quedará representada por los valores +1 o -1, según que sea positiva o negativa; cuando los

cambios no llegan a ese nivel, la correlación es imperfecta, positiva o negativa, oscilando entre 0 (correlación* nula) y 1, positivo o negativo (Figura 8).

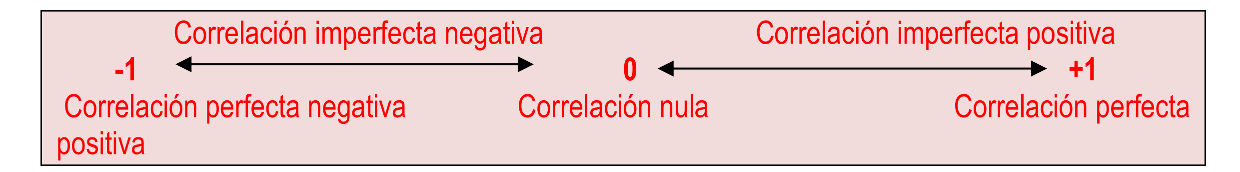


Figura 8: Valores posibles del coeficiente de correlación

En nuestro ámbito no cabe pensar en correlaciones perfectas, denominadas *funciones.* Así, la relación entre la longitud de la circunferencia –C- con la de su radio o su diámetro es una función, lo que nos permite conocer los valores de aquella a partir de los de estos:

$$C = 2\pi r = \pi d$$

2. Significación estadística* de un coeficiente de correlación*

Una cuestión importante es la de si una correlación solo es nula cuando su valor es, exactamente, 0. Y aquí se nos aparece de nuevo la Estadística inferencial.

Parece claro que si las dos series de datos abarcan todos los casos posibles y han sido medidas con instrumentos perfectos, una correlación* $r_{XY} = 0$ es una correlación nula. Sin embargo, la ciencia no utiliza todos los casos, bien sea por ser imposible, por ser muy caro, por no disponer de medios o porque -y esto es más importante- lo que se pretende es que lo descubierto en un caso pueda ser aplicado a otros de la misma naturaleza (por ejemplo: que la correlación encontrada este curso en niños de pre-escolar de 5 años pueda aplicarse a los de 5 años del curso siguiente).

Por ello, una pregunta aparentemente sencilla es: el valor r_{XY} encontrado en una muestra* ¿representa una auténtica correlación*? Técnicamente se dice: ¿Es estadísticamente significativo un valor de r_{XY}, por ejemplo de 0.12? Evidentemente 0.12 > 0 y parece que deberíamos afirmar que SI.

Sin embargo, la duda es inmediata: teniendo en cuenta que hemos obtenido los datos en unas series con solo algunos casos (muestras) y que los instrumentos de medida no son perfectos (tienen errores de medida debidos a las carencias en su fiabilidad), ¿podría ocurrir que tal valor no deba ser tomado en consideración (no sea estadísticamente significativo?. La respuesta es que SI; por ello, la Estadística nos ayudará a confiar o no en tal valor, a considerarlo como índice de una auténtica correlación o, por el contrario, como un valor que pudiera ser compatible con que, en el conjunto de casos (población), la correlación fuera nula.

Una cuestión fundamental al estudiar las correlaciones entre dos variables es la de si su magnitud nos permite pensar en una ***auténtica relación*** o si tal valor puede ser fruto del ***azar****, de la casualidad, en definitiva: ser casual o fortuito. En el primer caso afirmaremos que la correlación es ***estadísticamente significativa*** aunque asumimos cierto riesgo de error, concretado en un nivel de probabilidad* tan pequeño como decida el investigador. A este tema se le conoce como ***estimación de parámetros*.***

Así pues, analizaremos:

- *Qué es una correlación*
- *De qué tipo: positiva o negativa*
- *Qué intensidad tiene: perfecta o imperfecta, tanto positiva como negativa.*
- *Y dejaremos simplemente apuntada la idea de si es o no estadísticamente significativa, esto es, si la damos por tal o la consideramos fruto del azar* por los errores de muestreo y de medida.*

Por otra parte, conviene recordar los diferentes tipos de variables* y, en función de ellas, podemos establecer correlaciones entre:

- *Dos variables cualitativas: por ejemplo, sexo y grado universitario estudiado*
- *Dos variables ordinales: el puesto ocupado por un país en el Informe PISA y el rango y orden ocupado en analfabetismo.*
- *Dos variables cuantitativas discretas: pongamos por caso, las faltas de asistencia a clase y el curso académico que realizan los alumnos*
- *Dos variables cuantitativas continuas, como puede ser el de edad y talla.*
- *Cabe hablar de relación entre dos variables de diferente naturaleza: sexo e inteligencia, curso académico e inteligencia, raza y orden en la entrega de trabajos, …*

Dejemos constancia de que aquí solo pretendemos ejemplificar el concepto y tipos de correlación, acercándonos a su cálculo e interpretación, y que lo haremos con el coeficiente de correlación por excelencia, el de Pearson, representado, como hemos dicho, por r_{XY}, aplicable a la correlación entre variables cuantitativas medidas en escalas de razón y de intervalo.

Sobre los demás coeficientes de correlación (ordinal de Spearman, biserial, biserial por puntos, tetracórica o el coeficiente de asociación entre variables nominales) tendrán ocasión de acercarse a su conocimiento en el curso de la asignatura.

3. Aproximación al cálculo y representación gráfica

Con la simple finalidad de comprender lo que representa la correlación* presentaremos algún ejemplo sencillo; pongamos por caso, la correlación entre rendimiento académico, medido en un rango de 0 a 10 (variable A), y la inteligencia (variable B), medida con un test cuyo rango sea de 0 a 100 (tablas 8a y 8b). El cálculo no es objetivo de este curso introductorio.

He aquí los datos:

Variables	**PUNTUACIONES**												
X_A	7	4	3	2	7	8	9	6	4	8	4	5	6
X_B	45	34	40	58	70	70	88	63	45	56	41	47	69
XY	315	136	120	116	490	560	792	378	180	448	164	235	414
X^2	49	16	9	4	49	64	81	36	16	64	16	25	36
Y^2	2025	1156	1600	3364	4900	4900	7744	3969	2025	3136	1681	2209	4761

Variables	PUNTUACIONES (Continuación)							∑
X_A	6	7	9	1	3	6	7	**112**
X_B	62	80	94	24	15	45	73	**1119**
XY	372	560	846	24	45	270	511	**6976**
X^2	36	49	81	1	9	36	49	**726**
Y^2	3844	6400	8836	576	225	2025	5329	**60705**

Tabla 8 a: Puntuaciones en dos series de datos, A y B y cálculos para la obtención de r_{xy} (r_{AB})

Como se puede apreciar, la falta de ordenación hace difícil hacerse una idea sobre el tipo de relación. Sin embargo, la ordenación de una de las variables ya apunta hacia una relación imperfecta positiva:

X_A	1	2	3	3	4	4	4	5	6	6	6	6	7	7	7	7	8	8	9	9
X_B	24	58	15	40	34	41	45	47	45	62	63	69	45	70	73	80	56	70	88	94

Tabla 8 b: Puntuaciones en dos series de datos, A y B con valores de A ordenados

La representación gráfica, denominada *diagrama de dispersión*,* así lo confirma (figura 9):

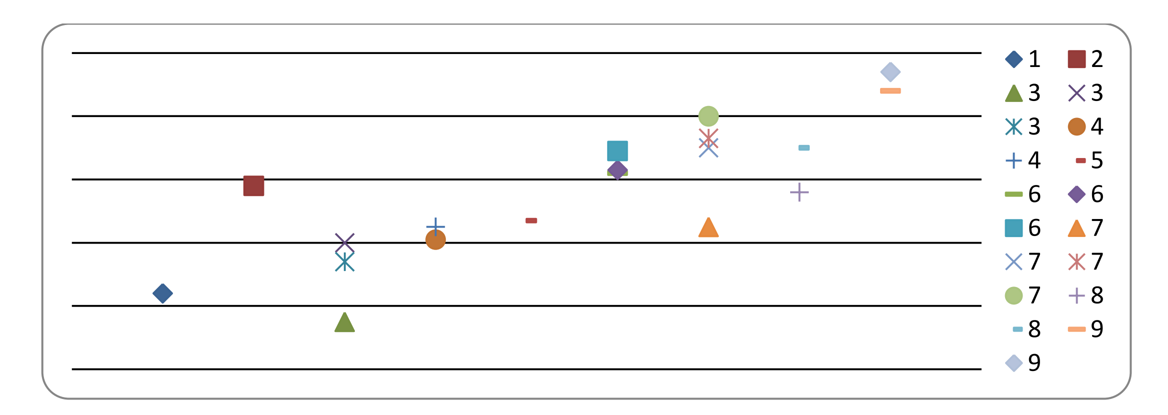


Figura 9: Diagrama de dispersión correspondiente a los datos de la tabla

Como se puede apreciar, los puntos reflejan la posición de cada sujeto en la serie A (que va de 0 a 10) y en la B, que va de 0 a 100. El punto en que se cruzan las líneas que van a los ejes de ordenadas y de abscisas representa a cada sujeto.

Es fácil comprender que la tendencia de las puntuaciones va de la parte inferior izquierda a la superior derecha (diagonal positiva) y que la correlación, siendo positiva, lo es imperfecta. La perfecta encontraría todos los puntos en la diagonal que fuera del 0-0 al 100-10.

La relación positiva apuntada se confirma aplicando la correspondiente ecuación a las parejas de datos (r_{XY}):

$$Ecuación\ \ 10: r_{XY} = \frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - (\sum X)^2\ .\ N\ \sum Y^2 - (\sum Y)^2}}$$

$$r_{XY} = \frac{20\ .6976 - 112\ .1119}{\sqrt{(20\ .726 - 112^2)\ .\ (20\ .70705 - 1119^2)}} = 0{,}79$$

La línea que mejor representa al conjunto de los puntos se denomina *recta de regresión;* cuando más se ajuste al conjunto de los puntos, mayor será la correlación; por otra parte, la inclinación de la pendiente también nos informa sobre la magnitud de la correlación. La *recta de regresión* sigue la ecuación *Y = a + bX*, donde Y es la puntuación en una de las dos variables a partir de la otra, X, siendo *a* y *b* dos constantes. La primera de ellas, *a*, es la ordenada en el origen y representa el valor de Y para X = 0. Estos aspectos no son objeto de estudio en el presente curso.

La relación entre dos variables se expresa, en términos estadísticos, mediante un ***coeficiente de correlación****. Sus valores pueden ir de -1 a +1; en tales casos, se habla de correlaciones perfectas. El valor 0 representa la correlación nula y los demás, correlación imperfecta, sea positiva o negativa.
La representación gráfica de dos pares de datos se conoce como ***diagrama de dispersión****, que nos permite apreciar de forma intuitiva el tipo de correlación y un acercamiento a su intensidad.
La línea que mejor representa al conjunto de pares de datos se conoce como ***recta de regresión***.

4. Interpretación

La interpretación de r_{XY} no es algo fácil ni definitivamente resuelto, salvo, claro está, en sus valores extremos, 0 y ±1.

Dos aspectos fundamentales deben ser tomados en consideración a la hora de interpretar los valores de r_{XY}:

a) Como ya hemos señalado previamente, si sus valores son o no estadísticamente significativos; esto resulta especialmente importante en el caso de valores bajos, próximo a 0, ya que bien podría ocurrir que una intensidad tan baja se debiera a factores como el azar*, que nada tienen que ver con una relación auténtica entre las dos variables correlacionadas.

   En el caso de ser significativos, como en todas las estimaciones por vía de inferencia, estamos aceptando una probabilidad* de error al afirmar que si o que no lo son.

b) Cuando un valor es estadísticamente significativo suele interesar graduar la intensidad de las correlaciones imperfectas, tanto positivas como negativas. Este punto no está definitivamente establecido, aunque hay algunas propuestas como la siguiente (tabla 9):

Magnitud de r_{XY}	Interpretación
0,00 a ± 0,20	Relación muy baja, despreciable
± 0,20 a ± 0,40	Relación baja
± 0,40 a ± 0,70	Relación sustancial
± 0,70 a ± 1,00	Relación alta o muy alta

Tabla 9. Acercamiento a la interpretación de los valores de r_{xy}

.No obstante, debemos indicar que la intensidad de r_{XY} varía en función de factores como el recorrido de las variables correlacionadas, el tamaño de la muestra (N), su variabilidad o dispersión y la fiabilidad de los instrumentos con los que se obtuvieron los valores de las variables. Con esto, queremos poner de relieve que un mismo valor de r_{XY} puede representar diferentes intensidades de correlación, lo que aconseja mucha prudencia a hora de interpretar este estadístico.

La interpretación de r_{XY} no es fácil pues sus valores están influidos por el tipo de variables correlacionadas, por el tamaño de la muestra, por su variabilidad o dispersión* y por las características técnicas de los instrumentos con los que hemos medido las variables*.
No obstante, lo primero y fundamental es saber si es estadísticamente significativa.

5. Principales aplicaciones

La gran utilidad de las correlaciones, como hemos reseñado, es la de ayudarnos a comprender la complejidad del ser humano al permitirnos conocer las relaciones existentes entre determinadas variables de su personalidad o de su actividad.

Pero conviene añadir dos muy importantes utilidades, a las que nos vamos a referir brevemente. En concreto son las que nos ayudan a establecer dos cualidades técnicas de gran relevancia que deben tener los instrumentos que utilizamos para recoger datos, la *fiabilidad* y la *validez.*

No es tarea de este curso meramente introductorio entrar con una mínima profundidad en ambos temas. Pero sí lo es, y con el ánimo de valorar su importancia, hacer saber que las técnicas estadísticas más utilizadas en uno y otro caso son las de *correlación**

5.1. Fiabilidad

Sin entrar en detalle, la fiabilidad de un instrumento nos informa del grado en que lo que mide lo hace con precisión, con el menor error de medida posible.

Pues bien: la técnica estadística utilizada es un *coeficiente de correlación** como el que hemos conocido, entre dos series de datos; dado que, como vamos a ver, ambas series se refieren a un mismo instrumento o a instrumentos equivalentes, la fiabilidad se representa por r_{XX}. Las series de datos son:

- *Las resultantes de dividir el instrumento en dos mitades (consistencia interna)*
- *Las surgidas de la aplicación por dos veces, debidamente separadas en el tiempo (estabilidad)*

- *Las obtenidas de la aplicación de dos instrumentos con las mismas características básicas (equivalencia).*

5.2. Validez

El gran problema de los instrumentos de medida en nuestro ámbito es el de la *validez,* definida generalmente como el grado en que miden lo que dicen medir. Aparentemente es una obviedad, pero en la realidad es una afirmación difícil mantener con pruebas.

Dos modalidades de validez se resuelven mediante coeficientes de correlación: la validez *concurrente* y la *predictiva.* En ambos casos, el coeficiente de validez se representa por r_{XY}.

La primera consiste en la correlación entre las puntuaciones obtenidas en el instrumento a validar –variable X- y en otros datos tomados como *criterio*, medidos por lo general de forma simultánea o muy próxima en el tiempo (variable Y). Por ejemplo: podemos intentar validar un instrumento para "medir" el grado de autoestima obteniendo la correlación entre las puntuaciones obtenidas en él por un grupo de alumnos –variable X- con las valoraciones realizadas sobre esta variable por un grupo de tutores que conocen bien a sus tutelados (criterio o variable Y), y recogidas ambas series de datos en tiempos muy próximos.

La segunda también utiliza el coeficiente de correlación, pero el *criterio* se mide pasado el tiempo para el que se desea predecir. Por ejemplo: si deseamos predecir a principio de curso la validez predictiva de una prueba diagnóstica de Estadística –variable X- en relación con las calificaciones de fin de curso, deberemos medir este criterio –calificaciones finales, variable Y- y obtener el coeficiente de correlación, denominado *predictivo.*

Si para la fiabilidad no suelen admitirse valores por debajo de 0.9, aquí no es fácil encontrar correlaciones superiores a 0.6 o 0.7, lo que hace que cualquier predicción implique asumir amplios márgenes de error.

De entre las aplicaciones o utilidades de la correlación destacamos las tres siguientes:
- Facilitar la interpretación de las relaciones entre variables.
- Calcular la fiabilidad de los instrumentos de medida (estabilidad, equivalencia, consistencia interna)
- Obtener indicios del grado de validez, predictiva y concurrente.

PARA SU REFLEXIÓN

A la vista de los contenidos de este capítulo, ¿considera que su actitud inicial hacia la asignatura ha mejorado? Si no es así, ¿a qué lo achaca?

¿Considera que el anterior contenido le es de utilidad?

A su juicio, comprender y ser capaz de utilizar los conocimientos del tema, ¿le serán de utilidad cuando comience a estudiar la asignatura?

A su juicio, comprender y ser capaz de utilizar los conocimientos del tema, ¿le serán de utilidad en su vida profesional?

Es probable que haya cuestiones del tema que no haya comprendido. Le invitamos a comunicarlas en el cuestionario de evaluación del curso.

CAPÍTULO 5. LA CURVA NORMAL DE PROBABILIDADES

Les recomiendo el visionado de la siguiente grabación:

http://www.intecca.uned.es/portalavip/grabacion.php?ID_Grabacion=61221&ID_Sala=65485&hashData=3badc56c38a8d519c2b816da0258141a&paramsToCheck=SURfR3JhYmFjaW9uLElEX1NhbGEs

La Real Academia de la Lengua define “modelo” como *arquetipo o punto de referencia para imitarlo o reproducirlo… Representación en pequeño de alguna cosa. Esquema teórico, generalmente en forma matemática de un sistema o de una realidad compleja…*

En su momento, definimos “modelo” como una representación de la realidad, una representación simplificada, ideal. Afirmábamos que la figura geométrica “cono”, es un modelo, como la esfera, la pirámide o el prisma; estamos ante modelos construidos por el hombre, pero que no están en la Naturaleza. Nadie ha visto en la realidad un cono pero sí objetos cónicos, y citábamos la cumbre del Teide.

Los “modelos” están en el día a día. Pensemos en la ropa de diferentes tallas. Cuando vamos a unos grandes almacenes encontramos ropa de todo tipo de diferentes tallas. Son modelos ideales; cuando nos probamos un traje puede que nos venga perfecto, pero es más común que resulte un poco más ancho, o largo, o estrecho; que tengan que sacarnos un cm., el dobladillo, o encoger la cintura, o subir el hombro… Estamos hablando de las “imperfecciones” de la mujer o del hombre real en relación con el modelo ideal como es la talla. Cuando la distancia entre el modelo y la persona real es muy grande es porque esa talla no representa el modelo al que se acomoda, y el dependiente le busca un traje de otra talla.

Quedémonos con esa idea de modelo como algo ideal, como una representación idealizada, simplificada, de la realidad. Pero seamos conscientes de que, gracias a ello, nos es posible acercarnos a la medida de la superficie de una montaña, o de su volumen. Gracias a ello, la industria textil puede hacer grandes tiradas de trajes, abaratando los costes. La alternativa es ir a un sastre o una modista para que nos haga una prenda a medida. Y gracias a ello, aunque en marcos diferentes, se prueban embarcaciones, coches, aeronaves…

1. El modelo

Pues bien, la curva normal de probabilidades* es un modelo de gran utilización en nuestro ámbito de trabajo debido a Carl F. Gauss (1777-1885).

En Estadística, uno de los modelos más utilizados es el denominado “normal”, representado en la figura 10. No nos detendremos en los demás porque lo que nos interesa es comprender su sentido, uso y utilidad, y esto vale para otros modelos, como t, F o el ya citado χ^2.

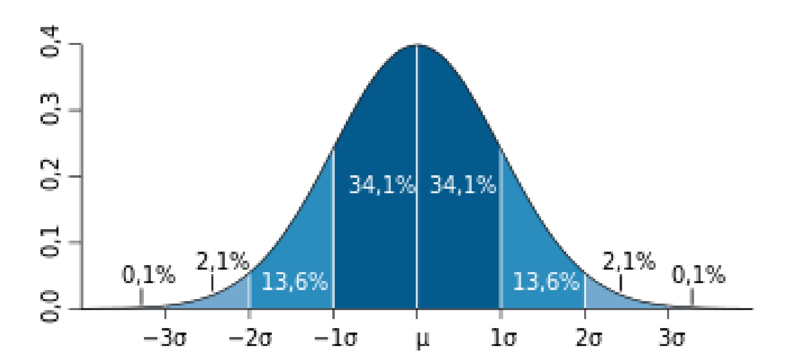


Figura10: Representación de la curva normal de probabilidades o campana de Gauss

Sobre unas coordenadas cartesianas, encontramos en el eje de abscisas diferentes valores típicos, expresados en σ alrededor del parámetro* media (μ), situado en el centro de la distribución, con la ordenada más elevada. Entre cada dos valores, la figura nos informa del % de casos que se dan en el modelo ideal, lo que, como veremos, tiene aplicaciones notables para interpretar las puntuaciones individuales de los sujetos.

Para interpretar valores que no se recogen en la figura tendríamos que acudir a la ecuación que rige el modelo (que aparece en la figura 5), pero no es necesario ya que disponemos de tablas estadísticas que nos permiten encontrar esos valores sin más esfuerzo que buscarlos en ellas (ver Anexo).

La curva normal de probabilidades* nos permite saber el % de casos que se deben encontrar entre dos valores típicos cualesquiera de una distribución empírica de datos que sigan ese modelo. Para averiguarlos debemos acudir a las tablas estadísticas, sin necesidad de hacer operaciones complejas.

2. Características

Las características fundamentales de este modelo* son:

- *El valor máximo de la serie se corresponde con la ordenada de la media. Expresando los valores en términos de puntuaciones típicas (z_i) a la media aritmética*le corresponde $z_i = 0$.*
- *La curva es simétrica con respecto a la ordenada de la media; por tanto, a ambos lados de la misma encontramos el 50 % de los casos de la distribución.*
- *Los valores de Media, Mediana y Moda coinciden.*
- *La curva disminuye progresivamente desde la ordenada de la media hacia ambos lados, encontrando sendos puntos de inflexión, a derecha e izquierda, que se corresponden con ± 1 σ.*
- *La curva es asintótica en relación con el eje de abscisas, esto es: eje y curva nunca llegan a cortarse o, de otro modo: solo se cortan en el infinito. Como consecuencia, la tabla de áreas de la curva normal nunca nos habla de sucesos seguros, cuya probabilidad* es 1; conforme nos alejemos hacia ambos lados, la probabilidad* se acercará a ese valor, pero nunca lo encontraremos.*

3. Principales aplicaciones

Una vez comprobado que una distribución empírica, real, se acomoda razonablemente al modelo*, podemos aplicar las características de este a aquella. De tal aplicación se siguen algunas utilidades relevantes.

3.1. Interpretar puntuaciones individuales

Ya hemos aludido a que una de sus aplicaciones fundamentales es la de ayudarnos a interpretar las puntuaciones de un sujeto situándolo en el contexto del grupo del que forma parte. Gracias a ello podemos construir *baremos* que nos permitan interpretar una puntuación en inteligencia, en autoestima, en producto interior bruto, en tasas de natalidad, etc.

Si tomamos una buena muestra de niños de 9 años, asistentes a escuelas de Educación Primaria, les aplicamos una prueba de conocimientos, la valoramos según una regla de medida previamente definida, hacemos una distribución de frecuencias y comprobamos que se acomoda razonablemente al modelo normal, aunque, como veíamos en la figura 2.b haya determinadas "imperfecciones" (como las aludidas en una montaña o en una talla), construimos con tales puntuaciones un baremo que permita ser aplicado a nuevas muestras de niños de 9 años de Escuela Primaria.

Como vemos, para ello hacen falta dos condiciones:

a) Que la muestra sea "buena", algo que hemos definido como que sea *representativa,* esto es: que tenga tamaño suficiente (hay tablas que nos dicen cuál es) y que sea seleccionada por procedimientos imparciales (en esencia, aleatorios).
b) Que, aplicada una prueba de *bondad de ajuste*, la probabilidad* a nuestro favor de que acomode al modelo, de que sea compatible con él, sea tan elevada como deseemos. Recordemos que, por muy elevada que sea, nunca podremos hablar de certeza o seguridad ya que los fenómenos aleatorios no lo permiten.

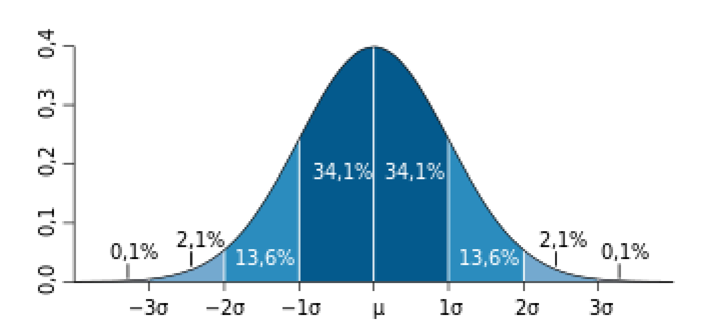
Con estas condiciones podemos construir el baremo en cuantiles, bien sean en cuartiles (Q), deciles (D) o centiles o percentiles (P). Para ello bastará con mirar en las tablas de la curva normal qué valor de z_i deja por debajo de sí el 25, el 50, el 60, el 75, el 80, el 83 % de los casos. En adelante, cualquier nuevo alumno cuya puntuación z_i sea como la de aquel que en el baremo le correspondió el percentil 83, o el 70, o el 35… la interpretaremos de este modo.

Se puede decir que las puntuaciones z_i no se encuentran en la realidad de una distribución empírica, sino que deben ser calculadas. Pero esto no es un problema ya que sabemos que $z_i = (X_i - \text{Media}) / s$.

Para que esté justificado interpretar las puntuaciones individuales tomando la curva normal* como referencia es preciso que la muestra en la que hemos obtenido las puntuaciones sea representativa, esto es, tenga suficiente tamaño y haya sido extraída por procedimientos imparciales (aleatorios) y que, aplicada la prueba de *bondad de ajuste**, los resultados nos informen de una alta probabilidad* a nuestro favor en el sentido de que ambas son compatibles.

Obviamente, relacionado con lo anterior, el modelo nos permite averiguar cuántos sujetos, o qué % de sujetos, quedan por encima de un valor típico, o entre dos valores de z_i (sea por encima o por debajo de la media, o uno por encima y otro por debajo.

Podemos apreciarlo de forma intuitiva volviendo a la figura 10:



Con ella, y consultando la tabla de áreas de la curva normal*, podemos comprobar que:

a) Que el 2,1 % de los casos se encuentran entre las puntuaciones directas a las que correspondan 2σ y 3σ; en efecto, en la tabla de áreas de la curva normal, por debajo de -2σ (área de la parte menor) queda el 2,28 %, mientras por debajo de -3σ queda el 0,13 %. Restando ambos valores tendremos 2,15 %.
b) Que el 13,6 % de los casos se encuentran entre -1σ y menos 2σ. La tabla de áreas de la curva normal nos indica que, por debajo de -1σ queda el 15,87 %; restando el 2,28 que hemos visto que corresponde a -2σ llegamos al 13,59 %
c) Que el 34,1 % se encuentra entre la ordenada de la media y menos –1σ. Como sabemos, la ordenada de la media aritmética deja por debajo de sí el 50 %; restando lo correspondiente a 1σ (15,87% de los caso) obtenemos el 34,13 % que figura en el gráfico.

Obviamente, para encontrar el % de casos que se encuentran entre ±1σ deberemos sumar los % correspondientes a la media menos 1σ (34,13) y a la media más 1σ (34,13); por tanto, estamos ante el 68,26 %.

Para transformar estos % a número de sujetos bastará en cada caso multiplicarlos por el valor de N en la distribución empírica de que se trate.

3.2. Atribuir probabilidades a los resultados del contraste de hipótesis*

La otra gran aplicación es la de atribuir probabilidades a determinados valores resultantes de las pruebas estadísticas relacionadas con el punto 6 del primer capítulo (*Poner a prueba diferentes formas de intervención sobre sujetos o grupos)*.

Veamos. Cuando ponemos a prueba dos métodos, por ejemplo, deseamos saber si los resultados nos permiten decidir si uno es mejor que otro, si da mejores resultados. El planteamiento del investigador, formulado como una *hipótesis**, sería:

Los resultados obtenidos con el método A son superiores a los logrados con el B

O dicho de otra forma más técnica:

Si los alumnos estudian con el método A, entonces obtendrán mejores resultados que si lo hacen con el B.

El planteamiento seguido en Estadística, lleno de prudencia, dado que queremos probar o contrastar esta hipótesis* “en general” y no solo para los dos grupos de alumnos, consiste en poder rechazar lo que conocemos como *hipótesis nula o de nulidad** (representada por H_0 frente a la del investigador, representada como H_1). Esta hipótesis dirá que no existen diferencias entre los resultados de ambos métodos, o que las que puedan existir pueden ser explicadas por efecto del azar*, de la casualidad, en una palabra: que son casuales o fortuitos, que no son estadísticamente significativos*.

Pensemos en dos métodos para el aprendizaje de los idiomas con niños de Primaria. Es evidente que si ambos dan, después de un período de prueba, la misma media aritmética, no tenemos razones para pensar que uno es mejor que el otro.

Pero lo normal es que una media aritmética sea superior a otra. ¿Podemos considerar que un método cuya media aritmética sea de 5,3 es mayor que una de 5.1? Desde luego, 5.3 > 5.1, pero, al igual que nos ocurría con el coeficiente de correlación, ¿no podría ocurrir que esta diferencia fuera casual, esto es: deberse al azar*?

Tengamos en cuenta que trabajamos no con todos los casos y que nuestras medidas no son tan perfectas como las utilizadas para medir la talla o el peso, que tienen errores de medida debidos a que su fiabilidad dista mucho de ser perfecta.

La segunda gran aplicación, propia de la Estadística inferencial, es la de decidir en un ***contraste de hipótesis**** si los resultados nos permiten afirmar que las diferencias son reales (estadísticamente significativas) o no, esto es: que puedan ser explicadas por el azar*, que surjan por pura casualidad.
Tal afirmación la haremos siempre no en términos de ***certeza*** o seguridad sino de ***probabilidad*,*** una probabilidad a nuestro favor tan grande como deseemos.

Determinadas pruebas estadísticas, como t de Student o F de Snedecor contrastan las medias aritméticas alcanzadas por los dos valores de la variable independiente* (métodos A y B en nuestro caso) con una estimación de lo que podría explicarse por puro azar*. Cuando la resultante de esta comparación (cociente) tienen unos resultados con alta probabilidad* de deberse al azar*, no se acepta la hipótesis* del investigador sino que se mantiene la denominada *nula o de nulidad** que, en definitiva, dice: las diferencias encontradas no son reales, tienen una alta probabilidad* de explicarse por puro azar*. En definitiva: se afirma que las diferencias encontradas *no son estadísticamente significativas.*

Esto es lo que se quedaría representado por la figura 11.a

En tales casos, el investigador debe ser cauto, prudente: si cabe una posibilidad razonable que esa diferencia en 5.3 y 5.1 pueda ser explicada por azar*, no la asume como verdadera sino que se la atribuye al azar*.



Figura 11.a Representación intuitiva de una contraste de hipótesis no significativo en términos estadísticos

El problema que se plantea, justamente, es el de saber a partir de qué magnitud de las diferencias pueden ser tomadas, con alta probabilidad* a nuestro favor, como difícilmente explicables por azar* y razonablemente asumibles como resultado del efecto diferencial de los métodos contrastados (figura 11.b).



Figura11.b Representación intuitiva de un contraste de hipótesis probablemente significativo en términos estadísticos

Si en la figura 11.a vemos intuitivamente que la parte $V_{experimental}$ (varianza producida por las diferencias entre métodos) es aproximadamente igual e incluso menor que V_{error} (varianza explicable por el azar), en la 11.b apreciamos que la primera es sensiblemente superior a la segunda. Por tanto, en el primer caso no cabe aceptar la hipótesis de que los métodos produzcan diferencias en la variable dependiente; pero en el segundo la decisión no la podemos tomar sin más, sino acudiendo a la Estadística inferencial, mediante un *contraste de hipótesis** que nos ayudará a decidir a partir de las probabilidades de que nuestro resultado pueda explicarlo el azar o la casualidad .

Pues bien, la curva normal de probabilidades* (y otros modelos, como t, F o χ^2) nos ayudan a asignar esas probabilidades. Científicamente se viene asumiendo que cuando las probabilidades a nuestro favor –*nivel de confianza*- son 95 o más % y, por tanto, las probabilidades de

equivocarnos son, como mucho, del 5 %, podemos aceptar la hipótesis* de que las diferencias se deben al método, aquí llamado variable independiente*, y no al puro azar*.

Las pruebas estadísticas utilizadas para contrastar hipótesis* siguen modelos diferentes. Sin embargo, en todos los casos el procedimiento se orienta en la misma dirección: decidir, con una alta probabilidad* a nuestro favor, si los resultados de la variable independiente* –método, técnica de motivación, sistema de disciplina…- sobre la variable dependiente*, se pueden atribuir a tales variables o pueden ser explicables como consecuencia del azar* o la casualidad, si son algo fortuito o real.

Conviene dejar constancia, no obstante, de que precisamente, la prueba F no sigue el modelo* normal (tiene sus propias tablas), pero todo el razonamiento utilizado le es aplicable.

PARA SU REFLEXIÓN

A la vista de los contenidos de este capítulo, ¿considera que su actitud inicial hacia la asignatura ha mejorado? Si no es así, ¿a qué lo achaca?

¿Considera que el anterior contenido le es de utilidad?

A su juicio, comprender y ser capaz de utilizar los conocimientos del tema, ¿le serán de utilidad cuando comience a estudiar la asignatura?

A su juicio, comprender y ser capaz de utilizar los conocimientos del tema, ¿le serán de utilidad en su vida profesional?

Es probable que haya cuestiones del tema que no haya comprendido. Le invitamos a comunicarlas en el cuestionario de evaluación del curso.

SEGUNDA PARTE
Pruebas de Autoevaluación

CAPÍTULO 1

La evaluación ha venido siendo una calificación emitida por el profesor que decide sobre el éxito o el fracaso del alumno en ciertos estudios. Sin embargo, pedagógicamente hablando, esa es solo una de sus funciones, siendo más importante la de comprobar el grado de aprendizaje que se va logrando, identificando los aspectos positivos y las carencias o limitaciones, a fin de tomar a tiempo las medidas correctores que sean precisas.
A esta evaluación, orientada a la mejora, se suele denominar *formativa* y puede realizarse por el propio alumnado con las orientaciones del profesorado.
Pues bien: en este curso 0 vamos a acudir a la autoevaluación, evaluación formativa realizada por los propios alumnos. Sobre los diversos temas se les plantearán cuestiones a las que ustedes deberán responder. Una vez cumplimentadas las respuestas a cada tema, podrán contrastarlas con las ofrecidas por el equipo docente y, como personas adultas, tomar las decisiones pertinentes: pasar de tema, corregir sus errores, eliminar sus carencias.
Como podrán apreciar, las respuestas ofrecidas por los profesores también les van a permitir afianzar su aprendizaje.

CAPÍTULO 1. CUESTIONES

1. INTERPRETAR PUNTUACIONES INDIVIDUALES

1.1. ¿Sabe interpretar el peso de una persona? ¿Qué elementos necesita para interpretar con precisión la altura o el peso de una persona?

1.2. ¿Sabría interpretar la puntuación en una prueba objetiva? ¿Qué elementos necesitaría para hacerlo correctamente?

1.3. En el caso de intentar medir la inteligencia, ¿cuál es la primera dificultad?. Salvada esta, ¿cómo podemos proceder para interpretar una puntuación individual?

2. CARACTERIZAR GRUPOS

2.1. ¿Cuántos tipos de medidas conoce para caracterizar estadísticamente a un grupo?

2.2. Un profesor desea conocer las características fundamentales de un grupo de alumnos en la asignatura que imparte. ¿Qué información puede proporcionarle la Estadística?

3. RAER INFORMACIÓN PARA LA TOMA DE DECISIONES

3.1. El científico utiliza la Estadística para hacer avanzar el saber. Entre los profesionales, la Estadística nos ayuda a:

3.2. Enumere algunas de las decisiones que se facilitan a partir de la aplicación de instrumentos de recogida de datos y de medida.

4. IDENTIFICAR RELACIONES ENTRE VARIABLES

4.1. Dos variables pueden variar sin relación entre sí. Pero también pueden variar conjuntamente, esto es: co-variar. En este caso afirmamos que las variables están…?

4.2. La relación entre dos variables puede ser:

4.3. Atendiendo a su intensidad, la relación o correlación puede ser:

4.4. Los siguientes valores representan diferentes tipos de correlación: -0.4; -0,9; 0; 0,4; 0.9. Denomínelos y ordénelos atendiendo a su intensidad.

4.5. Una correlación perfecta se representa por…?

4.6. Una aplicación de las correlaciones es la…:

4.7. En nuestros ámbitos, la predicción se utiliza para:

4.8. Para predecir debemos conocer las variables que habría que modificar a fin de evitar el cumplimiento de predicciones negativas. A su juicio, para predecir el éxito o el fracaso de un alumno en asignaturas científicas, ¿cuál es el mejor predictor de entre los dos siguientes, inteligencia y regularidad en el estudio?. Razone la respuesta.

4.9. ¿Y entre inteligencia y clase social?

5. APLICACIÓN DE LAS PROPIEDADES DE LOS MODELOS ESTADÍSTICOS

5.1. El conjunto de todos los casos o valores de un grupo en una característica concreta se denomina…? ¿Y el de un subconjunto del conjunto total?

5.2. ¿Cómo se denominan los valores medidos en un subconjunto que denominamos muestra? ¿Y sus correspondientes en la población?

5.3. El procedimiento para pasar de los valores de la muestra a la población corresponde a la Estadística que denominamos….?

5.4. Los valores obtenidos en las muestras son medidos. ¿Y los de la población?

5.5. Para estimar los parámetros aplicamos determinados modelos. ¿Es lícito aplicar un modelo a los datos obtenidos en una muestra?

5.6. ¿Cómo podemos decidir si unos datos empíricos se acomodan *razonablemente* a un modelo?

5.7. Cuando aplicamos una prueba de “bondad de ajuste”, ¿estamos seguros de que la distribución empírica sigue al modelo?

6. PONER A PRUEBA DIFERENTES FORMAS DE INTERVENCIÓN

6.1. ¿Cómo podemos decidir, con rigor y prudencia, si una forma de intervención –psicológica, pedagógica, social, médica…- es más eficaz que otra para alcanzar determinados logros o resultados?. (Entendamos por “intervención” una actuación del profesional, como una sesión de terapia, una reunión con el orientador, la utilización de un método, una forma de motivación …)

6.2. ¿Cómo se denomina el procedimiento por el que extendemos los resultados anteriores, obtenidos en un conjunto denominado muestra a todo el conjunto del que se ha obtenido la muestra? ¿Cómo se denomina este segundo conjunto?. La generalización* que realizamos mediante procesos de inferencia, ¿da lugar a resultados seguros?

6.3. ¿Cómo puede interpretar las puntuaciones individuales o grupales?

6.4. ¿Cuál es el modelo más habitual y conocido para interpretar puntuaciones de grupo?

6.5. ¿Cuándo es lícito utilizar un modelo, como el normal, para interpretar puntuaciones individuales y de grupo?

CAPÍTULO 1. RESPUESTAS

1. INTERPRETAR PUNTUACIONES INDIVIDUALES

1.1. ¿Sabe interpretar el peso de una persona? ¿Qué elementos necesita para interpretar con precisión la altura o el peso de una persona?

Es una cuestión por lo general, fácil. Pero lo es porque tenemos una idea aproximada de lo que es mucho, poco o normal.

Nos ayuda a la interpretación conocer si se trata de niños, jóvenes o adultos, si son niños o niñas, si son de un país conocido o desconocido (la etnia puede cambiar la interpretación…)

1.2. ¿Sabría interpretar la puntuación en una prueba objetiva? ¿Qué elementos necesitaría para hacerlo correctamente?

Por lo general, estamos acostumbrados a interpretar el resultado en una escala de 0 a 10, sabiendo que el 0 significa la peor de las calificaciones y el 10 la mejor; entre uno y otro extremo nos hacemos una idea aproximada.

Sin embargo, una prueba objetiva puede tener valores inferiores a 0 y superiores o muy superiores a 10, y esto dificulta la interpretación.

Conocer cuál es el "suelo", o puntuación mínima, y el techo o puntuación máxima, nos ayuda a interpretar. Como nos ayuda conocer la ***unidad***, esto es, cuánto vale cada respuesta, si el valor es o no uniforme, si se resta o no por respuestas erróneas…

Y, desde luego, nos ayuda saber dónde se sitúa la "raya" de lo normal o, si se desea, del aprobado; como nos ayudará saber a partir de qué puntuación se asigna un notable o un sobresaliente.

1.3. En el caso de intentar medir la inteligencia, ¿cuál es la primera dificultad?. Salvada esta, ¿cómo podemos proceder para interpretar una puntuación individual?

La primera y principal dificultad es la de definir qué entendemos por "inteligencia". Carecemos de una definición única que nos permita acercarnos a su "medida".

Una vez definida, debemos elaborar reactivos –ítems- que, según tal definición, representen manifestaciones de "inteligencia".

Además, debemos decidir la regla de medida: cuánto "vale" cada respuesta correcta según la definición.

Y, por último, interpretar la puntuación atendiendo a los valores que se consideran propios del grupo al que pertenece cada persona; a estos valores suele denominársele "baremos". Una misma puntuación puede significar poco, bastante o mucha según se trate de niños o de adolescentes, de estudiantes universitarios de carreras científicas o tecnológicas o de humanidades, de adultos sin estudios o con estudios de diferentes niveles…

Con un baremo podemos situar a cada persona en su grupo, entre los 10 mejores, entre los "normales", entre el 10 % peores …

También nos puede ayudar a la interpretación conocer cuál es la puntuación representativa del grupo (media, mediana, moda) o el grado de dispersión (puntuaciones más o menos homogéneas, concentradas en el centro o en los extremos…)

2. CARACTERIZAR GRUPOS

2.1. ¿Cuántos tipos de medidas conoce para caracterizar estadísticamente a un grupo?

Se utilizan las medidas de *posición*, también denominadas de tendencia central precisamente porque suelen tener valores próximos a los centrales de la distribución de puntuaciones; de *dispersión*, porque nos informan si las puntuaciones individuales de los miembros del grupo se acumulan en la parte central, media o extrema de la distribución; en el primer caso, la dispersión o variabilidad es baja y se acerca a 0; en los otros dos se aparta progresivamente de 0; las de *forma* son de *asimetría* y de *apuntamiento,* y nos hacen saber si las dos partes de la distribución –inferior y superior a la media- son más o menos simétricas (asimetría) y si las puntuaciones del centro son más o menos apuntadas o achatadas que las de un determinado modelo, la curva normal.

2.2. Un profesor desea conocer las características fundamentales de un grupo de alumnos en la asignatura que imparte. ¿Qué información puede proporcionarle la Estadística?

Fundamentalmente le puede ofrecer tres tipos de medidas: *posición, dispersión y forma.* Con ellas puede conocer lo que es más frecuente o normal (posición), el grado de variabilidad de las puntuaciones individuales alrededor de las anteriores medidas (dispersión o variabilidad) y la forma que toma la distribución en relación con la simetría / asimetría y el apuntamiento / achatamiento.

2. EXTRAER INFORMACIÓN PARA LA TOMA DE DECISIONES

2.1. El científico utiliza la Estadística para hacer avanzar el saber. Entre los profesionales, la Estadística nos ayuda a:

Conocer nuestro campo –la Educación, la persona, la sociedad, la economía, la enfermedad…- para tomar decisiones bien respaldadas en ese ámbito.

2.2. Enumere algunas de las decisiones que se facilitan a partir de la aplicación de instrumentos de recogida de datos y de medida.

Diagnosticar, predecir, orientar, prevenir…

3. IDENTIFICAR RELACIONES ENTRE VARIABLES

3.1. Dos variables pueden variar sin relación entre sí. Pero también pueden variar conjuntamente, esto es: co-variar. En este caso afirmamos que las variables están…?

Relacionadas o correlacionadas.

3.2. La relación entre dos variables puede ser:

Positiva o negativa.

3.3. Atendiendo a su intensidad, la relación o correlación puede ser:

Perfecta e imperfecta. Si no se da correlación se dice que esta es nula.

3.4. Los siguientes valores representan diferentes tipos de correlación: -0.4; -0,9; 0; 0,4; 0.9. Denomínelos y ordénelos atendiendo a su intensidad.

-0.4 y -0.9 indican correlación negativa e imperfecta
0.4 y 0.9 representan correlación positiva e imperfecta
0 es la representación de una correlación nula.
-0.9 y 0.9 son las correlaciones más elevadas; -0.4 y 0.4, les siguen en intensidad; 0 es la correlación más baja.

3.5. Una correlación perfecta se representa por…?

$r_{xy} = -1$ o $r_{xy} = +1$; en estos casos estamos ante una función: conocido el valor en una variable podemos conocer de antemano el valor que alcanzará la otra variable.

3.6. Una aplicación de las correlaciones es la…:

Predicción, si bien con márgenes de error.

La relación entre la longitud de la circunferencia y su diámetro es una función:

$$C = 2\pi r$$

Conocido el valor de r, obtenemos el valor de C al multiplicarlo por 2π.

Sin embargo, en $r_{xy} = 0,7$, siendo X = inteligencia general e Y = resultados del aprendizaje en Física, conociendo el valor en X solo podemos *estimar* el de Y con un margen de error, margen que nos ayudará la Estadística a determinar con cierta probabilidad.

3.7. En nuestros ámbitos, la predicción se utiliza para:

Ayudarnos a tomar por anticipado las medidas que eviten que se cumplan los pronósticos cuanto estos son negativos.

3.8. Para predecir debemos conocer las variables que habría que modificar a fin de evitar el cumplimiento de predicciones negativas. A su juicio, para predecir el éxito o el fracaso de un alumno en asignaturas científicas, ¿cuál es el mejor predictor de entre los dos siguientes, inteligencia y regularidad en el estudio?. Razone la respuesta.

Horas de estudio, porque está en manos del alumno su más inmediata aplicación

3.9. ¿Y entre inteligencia y clase social?

La inteligencia. Es difícil de modificar en el corto plazo, pero es modificable; la clase social también es modificable pero, cuando una familia pueda mejorarla es posible que los hijos hayan acabado sus estudios.

4. APLICACIÓN DE LAS PROPIEDADES DE LOS MODELOS ESTADÍSTICOS

4.1. El conjunto de todos los casos o valores de un grupo en una característica concreta se denomina… ? ¿Y el de un subconjunto del conjunto total?

Población. Muestra

4.2. ¿Cómo se denominan los valores medidos en un subconjunto que denominamos muestra? ¿Y sus correspondientes en la población?¿Con qué tipo de letras se representan unos y otros.

Estadísticos. Parámetros. Los primeros se representan con letras latinas; los segundos, con letras griegas.

4.3. El procedimiento para pasar de los valores de la muestra a la población corresponde a la Estadística que denominamos….?

Inferencial.

4.4. Los valores obtenidos en las muestras son medidos. ¿Y los de la población?

Se trata de valores *estimados*; por tanto, cabe esperar alguna diferencia entre los estadísticos medidos y sus correspondientes parámetros estimados.

4.5. Para estimar los parámetros aplicamos determinados modelos. ¿Es lícito aplicar un modelo a los datos obtenidos en una muestra?

Lo es siempre que los datos reales o empíricos se acomoden razonablemente al modelo.

4.6. ¿Cómo podemos decidir si unos datos empíricos se acomodan *razonablemente* a un modelo?

Aplicando pruebas que denominamos de "bondad de ajuste". Una de ellas es la denominada χ^2 –ji o chi cuadrado- que nos indica si una distribución de datos empíricos se acomoda a un modelo, por ejemplo al modelo normal.

4.7. Cuando aplicamos una prueba de "bondad de ajuste" que nos indica que se ese ajuste, ¿estamos seguros de que la distribución empírica sigue al modelo?

NO. En Estadística inferencial siempre se habla en términos de probabilidad: probabilidad de acertar en nuestra decisión, asumiendo que esta pueda ser errónea. Lo que procuramos es que las probabilidades a favor sean muy elevadas, y las en contra muy limitadas. Estas probabilidades las establece previamente el investigador.
Cuando un jugador compra muchos números a la Lotería, las probabilidades de que le toque son más elevadas que las de otro jugador que solo participa con un número. Sin embargo, después puede ocurrir que la lotería le toque a este y no al primero. *Mutatis mutandis*, esta es la cuestión de la probabilidad.

5. PONER A PRUEBA DIFERENTES FORMAS DE INTERVENCIÓN

5.1. ¿Cómo podemos decidir, con rigor y prudencia, si una forma de intervención –psicológica, pedagógica, social, médica…- es más eficaz que otra para alcanzar determinados logros o resultados?. (Entendamos por "intervención" una actuación del profesional, como una sesión de terapia, una reunión con el orientador, la utilización de un método, una forma de motivación …)

Utilizando diseños experimentales, en los que se contrasta el efecto de esas intervenciones –variables independientes- sobre la variable dependiente –logros o resultados deseados- en condiciones rigurosas de control, a fin de evitar, con niveles de probabilidad tan elevados como deseemos, interpretaciones alternativas como el azar.
Llamamos variable independiente (V.I.) a aquella que maneja el investigador (la forma de intervención) y variable dependiente (V.D.) aquella en la que deseamos ver los efectos de la primera. Así, una V.I, pueden ser tres formas diferentes de motivar, o dos terapias alternativas, o dos formas de participación… y la V.D., los resultados académicos, la mejora del autoncepto o de las habilidades sociales.

5.2. ¿Cómo se denomina el procedimiento por el que extendemos los resultados anteriores, obtenidos en un conjunto denominado muestra, a todo el conjunto del que se ha obtenido la muestra? ¿Cómo se denomina este segundo conjunto?. La generalización* que realizamos mediante tales procesos, ¿da lugar a resultados seguros?

Este segundo conjunto se denomina *población*. Y el procedimiento de generalización o extensión de los resultados, *inferencia.*

La inferencia siempre se expresa en términos de *probabilidad*, por tanto, admitiendo determinados niveles de error en sus afirmaciones. Es el investigador el que decide de antemano cuáles son esos márgenes y la Estadística le ayuda a conocer los correspondientes resultados.

5.3. ¿Cómo puede interpretar las puntuaciones individuales o grupales?

No es una cuestión fácil, en particular las de dispersión y forma.

Sobre la interpretación de las puntuaciones individuales ya hemos señalado la posibilidad de compararlas con las medidas de posición y de dispersión, o la comparación con un baremo con diversos tipos de *cuantiles* (Cuartiles –Q-, deciles –D- y centiles o percentiles –P)

Para otras interpretaciones, como las puntuaciones de forma, se recurre con frecuencia a “modelos”, esto es, a distribuciones ideales a las que tienden las distribuciones reales y que se toman como referencia.

Cuando una distribución empírica o real se acomoda razonablemente a un modelo, podemos interpretar las puntuaciones de un grupo –y de un sujeto- en el marco de ese modelo.

5.4. ¿Cuál es el modelo más habitual y conocido para interpretar puntuaciones de grupo?

Sin duda, el normal o gaussiano: la denominada curva normal de probabilidades. Sin embargo, debe quedar claro que hay modelos alternativos que son objeto de estudio de la Estadística.

5.5. ¿Cuándo es lícito utilizar un modelo, como el normal, para interpretar puntuaciones individuales y de grupo?

Cuando determinadas pruebas estadísticas especialmente adecuadas, por ejemplo la de χ^2 -pruebas denominadas de “bondad de ajuste”- nos informan de que la probabilidad de que los datos reales o empíricos sean compatibles con el modelo es tan elevada como desee el investigador.

CAPÍTULO 2

CAPÍTULO 2. CUESTIONES

1. Los enunciados, cuestiones o preguntas en un test, en un cuestionario, por lo general reciben el nombre de:
2. La puntuación directa de un sujeto en un test, prueba objetiva, cuestionario… suele representarse por_____ y se lee:
3. Los números son el resultado de:
4. Los números en sentido pleno, estricto, son el resultados de comparar:
5. Los números más perfectos, como los que se utilizan para medir, pesar o contar la longitud, el peso o la edad, son considerados como propios de una escala de medida denominada:
6. Aquellos números como los anteriores, pero en los que el 0 es arbitrario, como ocurre con la temperatura, son propios de una escala de medida:
7. Aquellos números cuya diferencia entre dos valores consecutivos no es constante, y que únicamente indican posición, rango, orden, forman las escalas de medida
8. Cuando atribuimos números a objetos que no representan cantidad sino igualdad o desigualdad, estamos ante las escalas de medida
9. Cuando estamos ante objetos que no son directamente observables, como la seguridad en sí mismo, la clase social subjetiva, la dificultad subjetiva de una prueba…, la primera actividad para proceder a su medida es:
10. Una vez definido el objeto debemos encontrar ____________ o elaborar __________:
11. Acto seguido nos encontraremos con otro problema técnico para poder asignar los números. Se trata de elabora la:
12. Las escalas ordinales solo indican orden, pero es cierto que el orden también indica cantidad. Por ello, hay autores que a estas escalas las denominan:
13. ¿La temperatura correspondiente a 20 grados centígrados es el doble de la que corresponde a 10 grados? ¿Por qué?
14. La diferencia entre 20 y 30 grados centígrados, ¿es el doble de la que se da entre 10 y 15 grados centígrados? ¿Por qué?
15. Clasifique las siguientes variables atendiendo a la escala de medida y a su carácter cuantitativo o cualitativo: edad, sexo, número de alumnos, calificaciones en un examen, dureza de los cuerpos, clase social, grado universitario.
16. Interprete las siguientes puntuaciones individuales: Q_2; D_6 ; P_{81} ; x_i= 1; CI = 98; z = 0.
17. Ante una serie de valores numéricos resultantes de aplicar la regla de medida, la Estadística recomienda realizar como primera operación su__________, lo que permite apreciar:
18. Una distribución de datos con muchos valores puede presentarse de forma reducida mediante lo que se conoce como:

19. En una distribución de frecuencias, dividimos su recorrido o rango, representando por ___, en un conjunto de:
20. La frecuencia se representa por el símbolo _____. Entendemos por frecuencia:
21. Cuando pasamos de una distribución original de puntuaciones directas (X_i) a una distribución de intervalos (I_i) se produce cierta ________________de los datos originales, aunque no le damos importancia porque las diferencias tienden a:
22. En una distribución por intervalos, cada intervalo se representa por lo que conocemos como:
23. El tratamiento de datos cualitativos (información de entrevistas, documentos…) tiene sus peculiaridades. Hay programas estadísticos que nos ayudan al tratamiento e interpretación de tales datos, como:
24. La expresión $\sum_{i=1}^{N} A$ debe leerse:
25. ¿Cuál de las dos siguientes series de datos tiene la Media aritmética más elevada? ¿Cuál es más heterogénea?

    A) 5, 4, 4, 4, 4, 4, 4, 4, 4, 3

    B) 8, 8, 7, 6, 5, 3, 2, 1, 0, 0
26. ¿Cuál es el recorrido (R) de las dos series anteriores?¿Cuáles son los límites exactos de la serie A (es decir, aquellos valores que formarían parte de la distribución aunque no fueran valores enteros)?
27. Le proponemos una sencilla distribución de frecuencias sobre la que se le formulan las siguientes preguntas:

    a) ¿Cuál es el valor de N?

    b) ¿Cuál es su R, expresado de las dos formas posibles?

    c) ¿Cuál es la amplitud de los intervalos?

    d) ¿En qué intervalo se sitúa la Mediana y la Moda?

    e) ¿Cuál es la marca de clase del primero y del último de los intervalos?

    f) Represente gráficamente los datos

    g) Indique si, de forma puramente intuitiva, tiende a la una distribución gaussiana. Razone su respuesta.

    DATOS:

X_i	f_i
31-35	3
26-30	12
21-25	14
16-20	9
11-15	6
6-10	1

CAPÍTULO 2. RESPUESTAS

1. Los enunciados, cuestiones o preguntas en un test, en un cuestionario, por lo general reciben el nombre de:

   Ítems

2. La puntuación directa de un sujeto en un test, prueba objetiva, cuestionario… suele representarse por_____ y se lee:

   X_i; puntuación directa o bruta del sujeto i.

3. Los números son el resultado de:

   Pesar, medir o contar objetos.

4. Los números en sentido pleno, estricto, son el resultados de comparar:

   Una cantidad con una unidad. El número es el resultado de tal comparación.

5. Los números más perfectos, como los que se utilizan para medir, pesar o contar la longitud, el peso o la edad, son considerados como propios de una escala de medida denominada:

   De cociente o de razón.

6. Aquellos números como los anteriores, pero en los que el 0 es arbitrario, como ocurre con la temperatura, son propios de una escala de medida:

   De intervalo.

7. Aquellos números cuya diferencia entre dos valores consecutivos no es constante, y que únicamente indican posición, rango, orden, forman las escalas de medida

   Ordinales

8. Cuando atribuimos números a objetos que no representan cantidad sino igualdad o desigualdad, estamos ante las escalas de medida

   Nominales.

9. Cuando estamos ante objetos que no son directamente observables, como la seguridad en sí mismo, la clase social subjetiva, la dificultad subjetiva de una prueba…, la primera actividad para proceder a su medida es:

Definirlos.

10. Una vez definido el objeto debemos encontrar ____________ o elaborar __________:

    Manifestaciones acordes con tal definición; o elaborar reactivos coherentes con tal definición.

11. Acto seguido nos encontraremos con otro problema técnico para poder asignar los números. Se trata de elabora la:

    Unidad de medida

12. Las escalas ordinales solo indican orden, pero es cierto que el orden también indica cantidad. Por ello, hay autores que a estas escalas las denominan:

    Cuasi-cuantitativas

13. ¿La temperatura correspondiente a 20 grados centígrados es el doble de la que corresponde a 10 grados? ¿Por qué?

    NO, porque el 0º no indica ausencia de temperatura.

14. La diferencia entre 20 y 30 grados centígrados, ¿es el doble de la que se da entre 10 y 15 grados centígrados? ¿Por qué?

    SI, porque la unidad "grado" es constante a lo largo de toda la escala.

15. Clasifique las siguientes variables atendiendo a la escala de medida y a su carácter cuantitativo o cualitativo: edad, sexo, número de alumnos, calificaciones en un examen, dureza de los cuerpos, clase social, grado universitario.

    Edad: Escala de razón. Cuantitativa continua
    Sexo: Escala nominal. Cualitativa, dicotómica
    Número de alumnos: Escala de razón. Cuantitativa discreta
    Dureza de los cuerpos: Escala ordinal. Cuasi-cuantitativa
    Clase social: Escala nominal. Cualitativa, politómica
    Grado universitario: Escala nominal. Cualitativa, politómica.

16. Interprete las siguientes puntuaciones individuales:

    Q_2= el sujeto se encuentra en el cuartil 2º, esto es, entre el 25 y el 50% inferior
    D_6 = el sujeto se encuentra en el decil 6º, esto es, supera al 60% del grupo
    P_{81} = el sujeto se encuentra en el centil o percentil 81, superando al 81 % del grupo
    x_i= 1; el sujeto i tiene una puntuación que supera la media en una unidad.
    CI = 98: el sujeto tiene un cociente intelectual un poco inferior al normal de su grupo (que sería 100). Por tanto, su Edad Mental (EM) es un poco inferior a su Edad Cronológica)

z = 0. El sujeto tiene una puntuación directa igual a la media del grupo.

17. Ante una serie de valores numéricos resultantes de aplicar la regla de medida, la Estadística recomienda realizar como primera operación su__________, lo que permite apreciar:

    Su ordenación creciente o decreciente. Permite conocer las puntuaciones extremas, la continuidad o no a lo largo de la distribución y la regularidad o no de la distribución a lo largo del recorrido de la variable.

18. Una distribución de datos con muchos valores puede presentarse de forma reducida mediante lo que se conoce como:

    Distribución de frecuencias.

19. En una distribución de frecuencias, dividimos su recorrido o rango, representando por ___, en un conjunto de:

    R; intervalos.

20. La frecuencia se representa por el símbolo _____. Entendemos por frecuencia:

    Representación: f_i (frecuencia del intervalo i). Frecuencia es el número de casos que se da en una puntuación o en un intervalo

21. Cuando pasamos de una distribución original de puntuaciones directas (X_i) a una distribución de intervalos (I_i) se produce cierta ________________ de los datos originales, aunque no le damos importancia porque las diferencias tienden a:

    Deformación. Compensarse.

22. En una distribución por intervalos, cada intervalo se representa por lo que conocemos como:

    Marca de clase, o punto medio del intervalo.

23. El tratamiento de datos cualitativos (información de entrevistas, documentos…) tiene sus peculiaridades. Hay programas estadísticos que nos ayudan al tratamiento e interpretación de tales datos, como:

    Atlas-ti o AQUAD.

24. La expresión $\sum_{i=1}^{N} A$ debe leerse:

    Súmense todos los aciertos desde el primero (1) al enésimo (N). Recordemos que hemos simbolizado por A las respuestas corresctas en una prueba objetiva.

25. ¿Cuál de las dos siguientes series de datos tiene la Media aritmética más elevada? ¿Cuál es más heterogénea?

    A) 5, 4, 4, 4, 4, 4, 4, 4, 4, 3

B) 8, 8, 7, 6, 5, 3, 2, 1, 0, 0

En ambos casos, la Media es 4; la más heterogénea es la serie B, que cuenta con puntuaciones bastante apartadas de la Media.

26. ¿Cuál es el recorrido (R) de las dos series anteriores?¿Cuáles son los límites exactos de la serie A (es decir, aquellos valores que formarían parte de la distribución aunque no fueran valores enteros)?

A): 5 - 3 + 1 = 3; B) 8 – 0 + 1 = 9; 2,5 y 5,5.

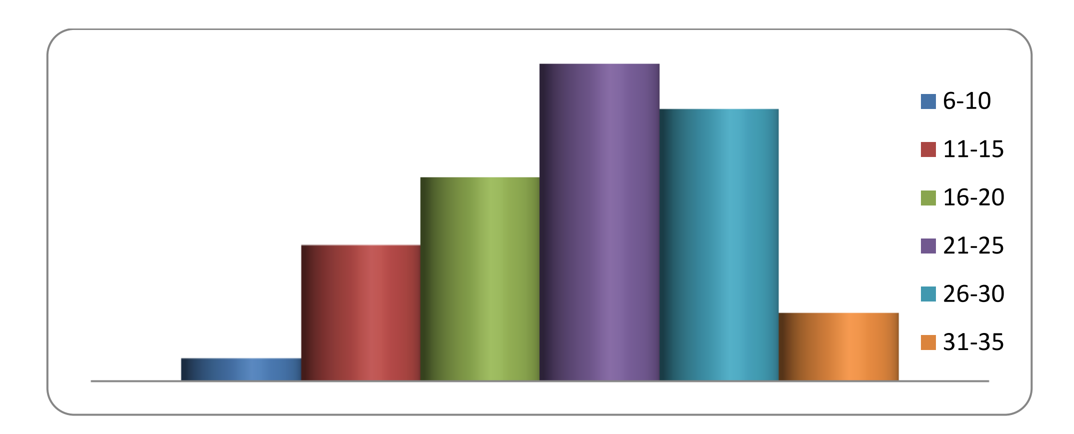
27. Le proponemos una sencilla distribución de frecuencias sobre la que se le formulan las siguientes preguntas:

a) ¿Cuál es el valor de N?
b) ¿Cuál es su R, expresado de las dos formas posibles?
c) ¿Cuál es la amplitud de los intervalos?
d) ¿En qué intervalo se sitúa la Mediana y la Moda?
e) ¿Cuál es la marca de clase del primero y del último de los intervalos?
f) Represente gráficamente los datos
g) Indique si, de forma puramente intuitiva, tiende a la una distribución gaussiana. Razone su respuesta.

DATOS:

X_i	f_i
31-35	3
26-30	12
21-25	14
16-20	9
11-15	6
6-10	1

a) 45
b) 35 – 6 + 1 = 30; 35,5 – 5,5 = 30
c) 5
d) 21 – 25
e) 8; 33.

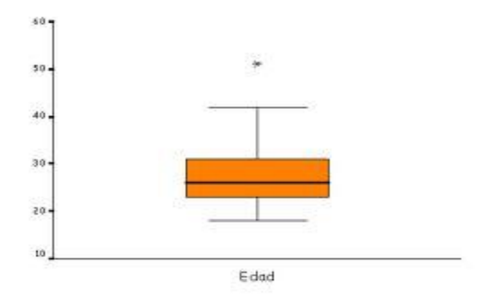


f) Si. Las frecuencias más elevadas se da en el centro de la distribución y disminuyen progresivamente hacia los extremos; mediana y moda se dan en el intervalo central y, previsiblemente, también la media; se aprecia cierta simetría.

CAPÍTULO 3

CAPÍTULO 3. CUESTIONES

1. **Las** ecuaciones que permiten calcular la media aritmética son: $\frac{\sum_{i=1}^{N} X_i}{N}$ y $\frac{\sum_{i=1}^{N} X_i f_i}{N}$. Indique cómo se leen y para qué casos se aplica cada una.
2. Para el cálculo de ciertas medidas de dispersión o variabilidad, como la desviación media con respecto a la media o la desviación típica debemos evitar resultados nulos. ¿Qué procedimiento debemos seguir en ambos casos?
3. En una de las medidas individuales, ligadas a la desviación típica, ¿qué representa un valor igual a 0? ¿A qué se debe? ¿Es una puntuación alta, media o baja?
4. En una curva normal ideal (el modelo) el valor que tiene la ordenada más elevada se corresponde con una puntuación directa igual a____ y con una puntuación z igual a_____:
5. Entre datos medidos y datos estimados, ¿cuáles vienen afectados por los errores ligados a la inferencia?
6. En el caso anterior, ¿cuáles creemos que representan mejor al parámetro?
7. Cuando se habla de estimar –inferencia- lo hacemos suponiendo que las muestras tienen calidad, esto es, que son representativas. ¿Qué cualidades hacen de una muestra una realidad representativa?
8. Para ayudarnos a decidir cuál de entre dos o más muestras presenta más dispersión o variabilidad utilizamos una medida conocida como:
9. El conjunto de puntuaciones situadas entre Q_1, y Q_3 representa el ___ % central de la serie y se denomina_____. Si ese valor se divide por 2, se conoce como:______
10. Para interpretar el grado de simetría / asimetría de una distribución tomamos como referencia:
11. Una distribución asimétrica, representada por $g_1 > 0$, la asimetría es _____ y se inclina hacia:
12. Una distribución leptocúrtica, representada por ___ tiene _____ apuntamiento que la distribución normal.
13. La distribución que presenta un apuntamiento igual o muy próximo al del modelo normal se conoce como:
14. Indique las aportaciones de la siguiente figura:



CAPÍTULO 3. RESPUESTAS

1. Las ecuaciones que permiten calcular la media aritmética son: $\frac{\sum_{i=1}^{N} X_i}{N}$ y $\frac{\sum_{i=1}^{N} X_i f_i}{N}$. Indique cómo se leen y para qué casos se aplica cada una.

   Súmense todas las puntuaciones directas (X_i) desde la primera (i) a la última (N, enésima) y divídase la suma por el número total de casos o puntuaciones (N).

   Súmense los productos de todas las puntuaciones directas (X_i) desde la primera (i) a la última (N, enésima) por su frecuencia (f_i) y divídase la suma por el número total de casos o puntuaciones (N).

2. Para el cálculo de ciertas medidas de dispersión o variabilidad, como la desviación media con respecto a la media o la desviación típica debemos evitar resultados nulos. ¿Qué procedimiento debemos seguir en ambos casos?

   En el caso de la desviación media, tomar las diferencias en valor absoluto; en la ecuación correspondiente esto queda representado por | |

   En el de la desviación típica, elevando las diferencias al cuadrado.

3. En una de las medidas individuales, ligadas a la desviación típica, ¿qué representa un valor igual a 0? ¿A qué se debe? ¿Es una puntuación alta, media o baja?

   Se trata de la medida conocida como z_i (puntación z del sujeto i). Para obtener un valor 0 su puntuación individual (X_i) debe ser igual a la Media aritmética. Por tanto representa a un sujeto que se encuentra justamente en el centro de la distribución: una puntuación media.

4. En una curva normal ideal (el modelo) el valor que tiene la ordenada más elevada se corresponde con una puntuación directa igual a_______ y con una puntuación z igual a_____:

   La Media aritmética del grupo. 0

5. Entre datos medidos y datos estimados, ¿cuáles vienen afectados por los errores ligados a la inferencia?

   Los valores estimados.

6. En el caso anterior, ¿cuáles creemos que representan mejor al parámetro?

   Los estimados

7. Cuando se habla de estimar –inferencia- lo hacemos suponiendo que las muestras tienen calidad, esto es, que son representativas. ¿Qué cualidades hacen de una muestra una realidad representativa?

Que esté seleccionada por un procedimiento imparcial, como es el caso de la selección aleatoria, y que su tamaño sea suficiente.

8. Para ayudarnos a decidir cuál de entre dos o más muestras presenta más dispersión o variabilidad utilizamos una medida conocida como:

   Coeficiente de variación, igual al cociente entre s y Media (desviación típica y media aritmética.

9. El conjunto de puntuaciones situadas entre Q_1, y Q_3 representa el ___ % central de la serie y se denomina_____. Si ese valor se divide por 2, se conoce como:______ *recorrido intercuartílico*, y su división por 2 *recorrido semi-intercuartílico.*

   50. Recorrido intercuartílico. Recorrido semi-intercuartílico.

10. Para interpretar el grado de simetría / asimetría de una distribución tomamos como referencia:

    La distribución normal o gaussiana

11. Una distribución asimétrica, representada por $g_1 > 0$, la asimetría es _____ y se inclina hacia:

    Positiva. La derecha

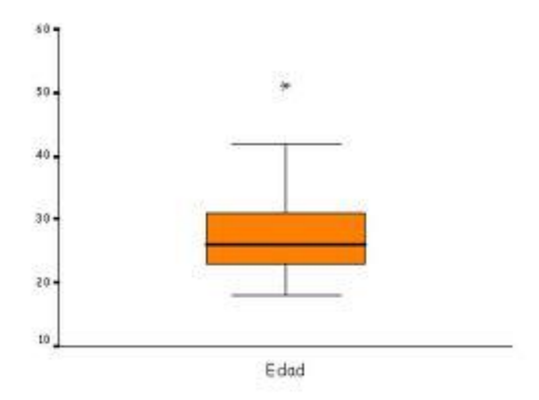
12. Una distribución leptocúrtica, representada por ___ tiene _____ apuntamiento que la distribución normal.

    $g_2 > 3$; mayor.

13. La distribución que presenta un apuntamiento igual o muy próximo al del modelo normal se conoce como:

    Mesocúrtica.

14. Indique las aportaciones de la siguiente figura:



Las puntuaciones superior e inferior, representadas por las dos líneas horizontales de los extremos

La Mediana, igual al Q_2 o cuartil 2, representada por la línea oscura dentro de la caja
El recorrido intercuartílico, representado por las dos líneas horizontales que limitan la caja.
La información contenida sobre este valor en el rectángulo central: en él podemos apreciar que la mediana queda en la parte inferior de la distribución y que los valores extremos se dan en mayor medida en la parte superior que en la inferior (bigotes).
Que la distribución sería asimétrica positiva.

CAPÍTULO 4

CAPÍTULO 4. CUESTIONES

1. En Estadística, la relación entre dos variables se denomina ___ y se simboliza, por lo general, como___
2. Cuando dos variables no están relacionadas estadísticamente se dice que son:
3. Cuando al variar los valores de una variable los de otra también lo hacen, se dice que_____, que las variables están ______
4. Si al aumentar el valor de una variable los de otra aumentan, la relación de denomina_____; en caso contrario la relación es_____
5. La intensidad de la relación oscila entre los valores de ____ y de ____
6. Los valores de r_{xy} que representan una covariación en la que al aumentar las puntuaciones en una variable disminuyen los de la otra, y que alcanzan valores > 0 y < 1 indican una correlación:
7. Cuando la relación entre dos variables es perfecta se denomina ____ y se representa por___.
8. En la conocida ecuación $C = 2\pi r$, el valor de la correlación existente es de:
9. La representación gráfica de la relación entre dos variables se conoce como:
10. Represente la relación entre las dos siguientes series de datos:

X	Y
1	22
2	20
3	19
4	25
5	30
6	17
7	32
7	40
9	37
10	39

11. En un diagrama de dispersión, los valores de la correlación imperfecta positiva tienden a ir del ángulo inferior ______ al superior________. En el caso anterior, por tanto, la correlación tenderá a ser:
12. Para que la correlación fuera perfecta negativa, los valores deberían:
13. La línea que mejor representa al conjunto de datos se conoce como:
14. Con los datos de la cuestión anterior, aplicando la ecuación 10, el valor de r_{XY} es:
15. ¿Puede ocurrir que un valor de $r_{XY} = 0.15$ no deba ser tomado en consideración?¿Qué procedimiento debe seguirse para tomar la correspondiente decisión?
16. ¿Qué queremos afirmar cuando decimos que un valor de r_{XY} es estadísticamente significativo?
17. La interpretación de r_{XY} es difícil. En algunas propuestas, un valor de r_{XY} de 0.30 sería interpretado como de una intensidad:
18. A la hora de interpretar valores significativos de r_{XY} se debe tener en cuenta que su intensidad varía en función de ciertos factores. Relacione los que conozca.

19. El denominado coeficiente de correlación de Pearson es adecuado para variables medidas en escalas de:
20. Con los mismos datos del problema anterior, pero alternando el orden de los valores de la variable Y, tal como aparecen a continuación, obtenga la representación gráfica y calcule el valor de r_{XY}. Razone los resultados.

X	Y
1	17
2	19
3	20
4	22
5	25
6	30
7	32
8	37
9	39
10	40

21. Dos de las aplicaciones más importantes del coeficiente de correlación son:

22. Hemos presentados tres tipos de fiabilidad, a saber:

23. Cuando se establece la correlación entre las puntuaciones obtenidas por un mismo grupo de sujetos en dos ocasiones diferentes, el coeficiente de fiabilidad se denomina:

24. La validez de un instrumento es un índice del grado en que este mide:

25. Hemos presentado dos tipos de validez:

26. Cuando se obtiene la correlación entre las puntuaciones obtenidas en el instrumento a validar y un ____________ medido prácticamente de forma simultánea, estamos ante la validez:

27. Los coeficientes de fiabilidad deben tener unos valores de:

28. Los coeficientes de validez no suelen superar los valores de:

CAPÍTULO 4. RESPUESTAS

1. En Estadística, la relación entre dos variables se denomina ___ y se simboliza, por lo general, como___

   Correlación. r_{XY}

2. Cuando dos variables no están relacionadas estadísticamente se dice que son:

   Independientes entre sí

3. Cuando al variar los valores de una variable los de otra también lo hacen, se dice que_____, que las variables están ______

   Co-varían, que están correlacionadas

4. Si al aumentar el valor de una variable los de otra aumentan, la relación de denomina_____; en caso contrario la relación es____

   Positiva; negativa.

5. La intensidad de la relación oscila entre los valores de ____ y de ____

   0 y 1

6. Los valores de r_{xy} que representan una covariación en la que al aumentar las puntuaciones en una variable disminuyen los de la otra, y que alcanzan valores > 0 y < 1 indican una correlación:

   Negativa e imperfecta

7. Cuando la relación entre dos variables es perfecta se denomina ____ y se representa por___.

   Función. 1

8. En la conocida ecuación $C = 2\pi r$, el valor de la correlación existente es de:

   + 1: perfecta positiva

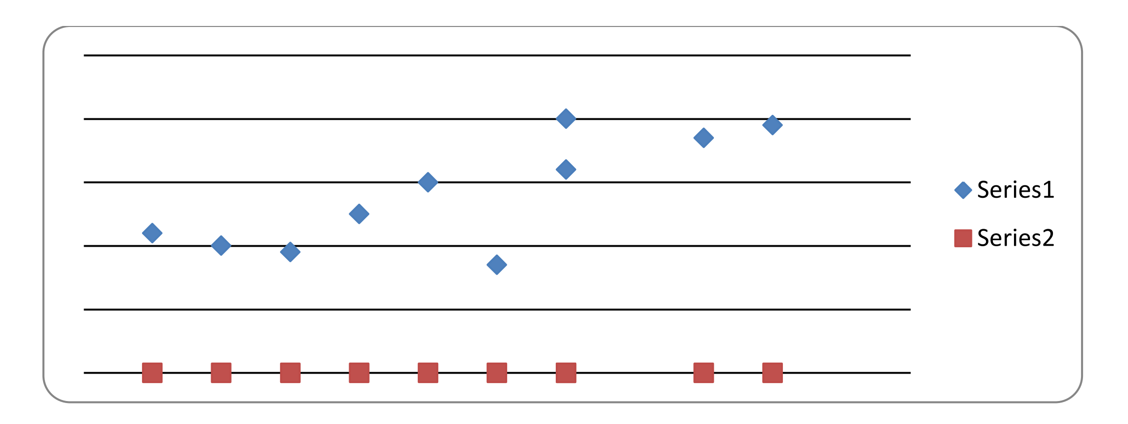
9. La representación gráfica de la relación entre dos variables se conoce como:

   Diagrama de dispersión

10. Represente la relación entre las dos siguientes series de datos:

X	Y
1	22
2	20
3	19
4	25
5	30
6	17
7	32
8	40
9	37
10	39

La representación correspondiente es el siguiente *diagrama de dispersión* realizado mediante Excell.



11. En un diagrama de dispersión, los valores de la correlación imperfecta positiva tienden a ir del ángulo inferior ______ al superior________. En el caso anterior, por tanto, la correlación tenderá a ser:

    Izquierdo, derecho, imperfecta positiva

12. Para que la correlación fuera perfecta negativa, los valores deberían ir:

    Del ángulo inferior derecho al superior izquierdo (compruebe que en el ángulo inferior derecho están las puntuaciones más elevadas en X y más bajas en Y, mientras que en el ángulos superior izquierdo ocurre todo lo contrario.

13. La línea que mejor representa al conjunto de datos se conoce como:

    Recta de regresión.

14. ¿Puede ocurrir que un valor de r_{XY} = 0.15 no deba ser tomado en consideración?¿Qué procedimiento debe seguirse para tomar la correspondiente decisión?

Efectivamente. Para tomar la decisión debemos comprobar si ese valor de r_{XY} es estadísticamente significativo (en el temario no hemos estudiado cómo se lleva a cabo)

15. Qué queremos afirmar cuando decimos que un valor de r_{XY} es estadísticamente significativo?

Que, con una elevada probabilidad a nuestro favor, decidida de antemano, ese valor se debe a la existencia de relación auténtica entre las variables y no al puro azar o casualidad.

16. La interpretación de r_{XY} es difícil. En algunas propuestas, un valor de r_{XY} de 0.30 sería interpretado como de una intensidad:

Baja

17. A la hora de interpretar valores significativos de r_{XY} se debe tener en cuenta que su intensidad varía en función de ciertos factores. Relacione los que conozca.

Recorrido de las variables, tamaño de la muestra (N), variabilidad o dispersión y fiabilidad de los instrumentos con los que se obtuvieron los valores de las variables.

18. El denominado coeficiente de correlación de Pearson es adecuado para variables medidas en escalas de:

Razón o cociente y de intervalo.

19. Con los datos de la cuestión 10, aplicando la ecuación 10, el valor de r_{XY} es:

Variables	Puntuaciones										∑
X	1	2	3	4	5	6	7	8	9	10	**55**
Y	22	30	19	25	30	17	32	40	37	39	**281**
XY	22	40	57	100	150	102	224	320	333	390	**1738**
X^2	1	4	9	16	25	36	49	64	81	100	**385**
Y^2	484	400	361	625	900	289	1024	1600	1369	1521	**8573**

$$\frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - (\sum X)^2 \, . \, N\sum Y^2 - (\sum Y)^2}}$$

$$r_{XY} = \frac{10\,.\,1738 - 55\,.\,281}{\sqrt{(10\,.\,385 - 55^2).\ (10\,.\,8573 - 281^2)}} = 0{,}81$$

Como puede apreciarse, una correlación positiva, imperfecta pero bastante elevada.

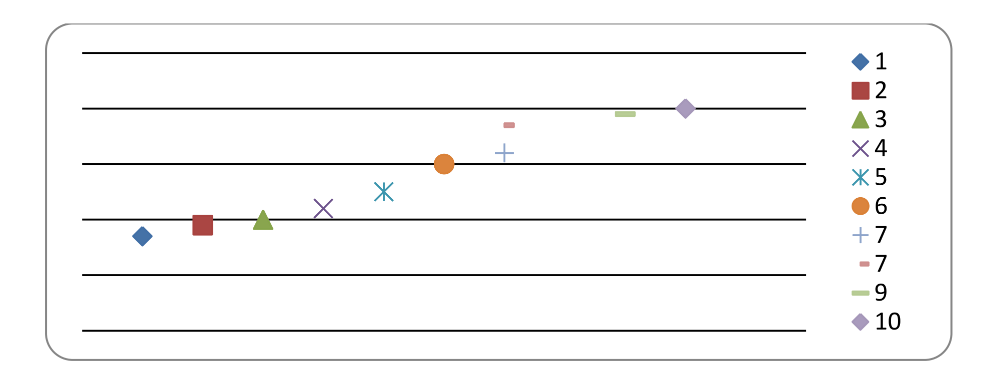
20. Con los mismos datos del problema anterior, pero alterando el orden de los valores de la variable Y, tal como aparecen a continuación, obtenga la representación gráfica y calcule el valore de r_{XY}. Razone los resultados.

X	Y
1	17
2	19
3	20
4	22
5	25
6	30
7	32
8	37
9	39
10	40

Variables	Puntuaciones										∑
X	1	2	3	4	5	6	7	8	9	10	**55**
Y	17	19	20	22	25	30	32	37	39	40	**281**
XY	17	38	60	88	125	180	224	296	351	400	**1779**
X^2	1	4	9	16	25	36	49	64	81	100	**385**
Y^2	289	361	400	484	625	900	1024	1369	1521	1600	**8573**

$$r_{XY} = \frac{10\,.\,1779 - 55\,.\,281}{\sqrt{(10\,.\,385 - 55^2).\ (10\,.\,8573 - \ 281^2\,)}} = 0{,}988$$

Su representación gráfica es la siguiente; en ella se puede apreciar un elevadísimo acercamiento a la diagonal que representa la correlación perfecta positiva. La diferencia con el caso anterior está en ∑XY; como puede comprobarse, el producto es más elevado, como consecuencia de que en la medida en que se elevan los valores en X también lo hacen en Y. Comparando los datos y los gráficos en el caso anterior, podemos ver que allí se daban algunas excepciones, el más llamativo es el del par 6 -17.



21. Dos de las aplicaciones más importantes del coeficiente de correlación son:

    El cálculo de la fiabilidad y de la validez de los instrumentos de recogida de datos y de medida.

22. Hemos presentados tres tipos de fiabilidad, a saber:

    Consistencia interna, estabilidad y equivalencia

23. Cuando se establece la correlación entre las puntuaciones obtenidas por un mismo grupo de sujetos en dos ocasiones diferentes, el coeficiente de fiabilidad se denomina:

    De estabilidad

24. La validez de un instrumento es un índice del grado en que este mide:

    Lo que dice o quiere medir.

25. Hemos presentado dos tipos de validez:

    Concurrente y predictiva.

26. Cuando se obtiene la correlación entre las puntuaciones obtenidas en el instrumento a validar y un ___________ medido prácticamente de forma simultánea, estamos ante la validez:

    Criterio, concurrente.

27. Los coeficientes de fiabilidad deben tener unos valores de:

    0.90 o superiores.

28. Los coeficientes de validez no suelen superar los valores de:

    0.60 / 0.70

CAPÍTULO 5

CAPÍTULO 5. CUESTIONES

1. Cuando representamos la realidad en forma simplificada e idealizada estamos acudiendo a un
2. En un modelo "normal", como la curva normal de probabilidades, el parámetro μ ocupa el lugar ______ y le corresponde la ordenada_____
3. En la curva normal de probabilidades, los valores de σ se sitúan en el eje de _____y sus valores extremos son:
4. Si un sujeto tiene $z_i = 0$ en una distribución compatible con el modelo normal, podemos saber que se trata de una persona con una puntuación directa
5. Cuando una distribución es compatible con el modelo normal, las tres medidas de posición
6. Una característica del modelo normal se refiere a los valores ±σ. Se trata de:
7. Los límites de una curva normal de probabilidades, esto es, los valores extremos de σ son________ ya que la curva es _________________ con respecto al eje de abscisas.
8. Para comprobar si una distribución de datos es compatible con la denominada curva normal, debemos aplicar una prueba de
9. Si una variable sigue el modelo normal, es fácil que nuestros datos empíricos lo corroboren si la muestra ha sido ___________________, esto es, si tiene un _________ suficiente y si los sujetos se han extraído por procedimientos _____________
10. El procedimiento más adecuado para extraer una muestra representativa es el:
11. Hemos hablado de dos grandes aplicaciones de la curva normal de probabilidades, en concreto de:
12. Un baremo es un regla que nos permite interpretar las puntuaciones directas de una distribución tomando como referencia.
13. Contrastamos hipótesis cuando sometemos a prueba, en condiciones de rigor (control de variables extrañas) el efecto de una o más variables independientes sobre la
14. Gracias a los diferentes modelos estadísticos, podemos atribuir probabilidades a nuestras hipótesis y decidir, con una determinada ____________ si son los efectos de la variable independiente los que producen los cambios constatados en la __________ o si estos pueden explicarse por puro_____
15. Cuando los resultados no nos permiten aceptar nuestra hipótesis debemos mantener la denominada hipótesis ________, según la cual las diferentes pueden ser explicadas como resultado:
16. Considera que el siguiente enunciado de una hipótesis, en una investigación destinada a contrastar diferencias a favor de uno de los dos métodos, es correcto?

Estudiar con dos métodos diferentes da lugar a mejorar los resultados

17. En una distribución empírica de N = 250 datos, compatible con la normal, con media = 30 y s = 10, ¿Cuántos sujetos se encontrarán por debajo de 30 puntos?
18. ¿Cuántos entre 20 y 30?
19. ¿Cuántos por debajo de 10?
20. ¿Cuántos por encima de 50?

CAPÍTULO 5. RESPUESTAS

1. Cuando representamos la realidad en forma simplificada e idealizada estamos acudiendo a un

   Modelo.

2. En un modelo “normal”, como la curva normal de probabilidades, el parámetro μ ocupa el lugar ______ y le corresponde la ordenada_____

   Central, más elevada.

3. En la curva normal de probabilidades, los valores de σ se sitúan en el eje de _____:

   Abscisas

4. Si un sujeto tiene z_i = 0 en una distribución compatible con el modelo normal podemos saber que se trata de una persona con una puntuación directa

   Igual a la medida aritmética

5. Cuando una distribución es compatible con el modelo normal, las tres medidas de posición

   Son iguales, coinciden

6. Una característica del modelo normal se refiere a los valores ±σ. Se trata de:

   Los puntos de inflexión

7. Los límites de una curva normal de probabilidades, esto es, los valores extremos de σ son________ ya que la curva es _________________ con respecto al eje de abscisas.

   Desconocidos; asintótica.

8. Para comprobar si una distribución de datos es compatible con la denominada curva normal, debemos aplicar una prueba de

   Bondad de ajuste

9. Si una variable sigue el modelo normal, es fácil que nuestros datos empíricos lo corroboren si la muestra ha sido ___________________, esto es, si tiene un _________ suficiente y si los sujetos se han extraído por procedimientos _____________

   Representativa; tamaño; imparciales.

10. El procedimiento más adecuado para extraer una muestra representativa es el:

    Aleatorio

11. Hemos hablado de dos grandes aplicaciones de la curva normal de probabilidades, en concreto de:

    La interpretación de las puntuaciones individuales y la atribución de responsabilidades a los resultados del contraste de hipótesis

12. Un baremo es un regla que nos permite interpretar las puntuaciones directas de una distribución tomando como referencia.

    La curva normal (u otro modelo, según los casos).

13. Contrastamos hipótesis cuando sometemos a prueba, en condiciones de rigor (control de variables extrañas) el efecto de una o más variables independientes sobre la

    Variable dependiente

14. Gracias a los diferentes modelos estadísticos, podemos atribuir probabilidades a nuestras hipótesis y decidir, con una determinada ____________ si son los efectos de la variable independiente los que producen los cambios constatados en la __________ o si estos pueden explicarse por puro_____

    Probabilidad; variable dependiente; azar o casualidad.

15. Cuando los resultados no nos permiten aceptar nuestra hipótesis debemos mantener la denominada hipótesis ________, según la cual las diferentes pueden ser explicadas como resultado

    Nula o de nulidad, del azar o de la casualidad.

16. Considera que el siguiente enunciado de una hipótesis, en una investigación destinada a contrastar diferencias a favor de uno de los dos métodos, es correcto?

    Estudiar con dos métodos diferentes da lugar a mejorar los resultados

    No lo es; el investigador no indica a favor de qué método deben producirse las diferencias.

17. En una distribución empírica de N = 250 datos, compatible con la normal, con media = 30 y s = 10, ¿Cuántos sujetos se encontrarán por debajo de 30 puntos?

    Sabiendo que la media es 30, y que la media divide la curva en dos partes simétricas, está claro que deberán situarse por debajo la mitad de 250: 125 casos.

18. ¿Cuántos entre 20 y 30?
    Las puntuaciones 20 y 30 equivalen a $\pm\ \sigma$; por tanto, entre ambos valores se encontrará, como hemos visto en el texto, el 34,26 %, esto es: aproximadamente 85 sujetos (34,26 de 250)

19. ¿Cuántos por debajo de 10?

La puntuación directa de 10 equivale a -2σ; por tanto, deja por debajo el 0.0228 %, o lo que es lo mismo, casi 6 sujetos (0,0228 por 254)

20. ¿Cuántos por encima de 50?

La puntuación 50 está 2σ por encima de la media aritmética; por lo tanto, estamos como en caso anterior, salvo que leemos los datos de forma diferente.



TERCERA PARTE
Glosario

CONCEPTOS

A lo largo de los materiales de aprendizaje van apareciendo conceptos cuya comprensión es fundamental para entender la temática.
En algún momento, el lector deseará acceder a los mismos; con la intención de facilitarle su lectura, estos conceptos aparecen en el texto con *. En tal caso, podrá venir a este documento y encontrar información sobre el citado concepto.

LISTADO DE CONCEPTOS

GLOSARIO

AZAR

Casualidad. Caso fortuito. Fenómeno que no sigue una regla, un orden, una ley conocida. En Estadística se contrastan las probabilidades a favor de las hipótesis del investigador en cuanto al efecto de las variables independientes sobre las dependientes contra la probabilidad de que los resultados sean debidos al azar, a la pura casualidad.

BAREMO

Tabla elaborada como regla para atribuir valor a las puntuaciones individuales o de grupo. Las puntuaciones del grupo inicial, una muestra que debe ser representativa de la población, sirven de regla para interpretar las puntuaciones de otros sujetos o grupos que reúnan las mismas características.

Es frecuente que a cada puntuación individual del grupo original se le asigne la correspondiente trasformación a puntuaciones como cuantiles (cuartiles, deciles o percentiles), puntuaciones típicas, C.I., etc. En adelante, a cualquier puntuación de un nuevo sujeto se le atribuye la correspondiente puntuación (cuantil…) del baremo.

DISPERSIÓN

Característica de un grupo que nos informa del grado en que las puntuaciones de los integrantes de un grupo se sitúan de forma más o menos cercana la medida de posición de que se trate (por ejemplo, de la media aritmética). Un grupo en que todos su miembros obtienen una puntuación igual a la medida de posición tiene una dispersión de 0; sin embargo, no existe un valor fijo de dispersión máxima.

Las medidas de dispersión o variabilidad más importantes y utilizadas son la desviación típica o la varianza.

CONTRASTE DE HIPÓTESIS

Proceso sistemático que concluye tomando una decisión sobre la hipótesis nula (H_0): rechazo o no rechazo, con la consiguiente aceptación o no de la hipótesis alternativa (H_1) asumiendo un riesgo de error tipo I $\leq \alpha$.

CONTROL

El método científico pretende establecer relaciones causales entre las variables relacionadas en su hipótesis. Lograr una meta tan elevada como este exige del investigador el dominio de la situación, de forma que, teniendo bajo su dominio la variable independiente, controle el conjunto de circunstancias, hechos, personas… que, además de dicha variable, puedan influir en la dependiente.

Si no fuera así, quedaría la duda de si la relación encontrada se debe a la variable independiente, a alguna de esas otras variables –convertidas en extrañas, esto es, en hipótesis rivales a la suya- o la la interacción entre unas y otras.

CORRELACIÓN

Entendemos por correlación la relación existente entre dos o más variables.

La correlación puede ser perfecta, positiva o negativa (valor de ± 1), nula (valor de 0), o imperfecta, que incluye toda la gama de valores que van de 0 a 1, tanto positivos como negativos. La correlación es positiva cuando los valores de las variables aumentan o disminuyen en la misma dirección, y negativa en caso contrario.

El índice de correlación –coeficiente de correlación- más conocido es el de Pearson, representado por r_{xy}.

CURVA NORMAL (DE PROBABILIDADES)

Se trata del modelo estadístico al que tienden con más frecuencia los datos empíricos. Cuando estos se acomodan razonablemente al modelo, podemos aplicarles sus propiedades.

Las características fundamentales de la *curva normal de probabilidades** son:

- *Tener como valor máximo el de la ordenada de la media*
- *Ser simétrica con respecto a la ordenada de la media.*
- *Presentar dos puntos de inflexión, uno para el valor de la Media más una desviación típica (media + s) y otro para la Media menos una desviación típica* (media – s)*
- *La curva es asintótica con respecto al eje de abscisas: por mucho que se prolongue a derecha e izquierda nunca llega a cortarlo. Por ello, nunca encontraremos una probabilidad = 1, que sería un suceso seguro.*

DESVIACIÓN TÍPICA

Medida de *dispersión* o *variabilidad.* Estadísticamente es la raíz cuadrada de la media de la suma de las desviaciones individuales de un grupo de sujetos, elevadas al cuadrado, con respecto a la media aritmética de un grupo.

La varianza es un índice del grado en que las puntuaciones individuales se agrupan más o menos en torno a la media del grupo; si todas las puntuaciones individuales coincidieran con la media, la varianza sería 0; cuanto más se aparten de ella, mayor valor alcanzará la varianza.

Esta medida es muy importante en la Estadística inferencial ya que se utiliza en las pruebas de contraste de hipótesis.

DIAGRAMA DE BARRAS

Representación gráfica especialmente adecuada a variables cualitativas; las barras, situadas unas a continuación de otras, tienen como base las diferentes categorías y como altura su frecuencia.

DIAGRAMA DE CAJA

Representación gráfica especialmente relevante por la gran cantidad de información que ofrece sobre un conjunto de puntuaciones originales. En concreto nos informa sobre:

- *Las puntuaciones extremas*
- *La Mediana, igual al Q_2 o cuartil 2*
- *El recorrido intercuartílico: Q_1 a Q3.*
- *La información contenida sobre este valor en el rectángulo central: en él podemos apreciar si se da equilibrio o no entre los diferentes cuartile.*
- *En ocasiones puede informarnos sobre las puntuaciones fuera de rango.*

DIAGRAMA DE DISPERSIÓN

Se trata de la representación gráfica de los pares de valores de un grupo de sujetos en dos variables. El diagrama nos ilustra de la existencia o no de correlación, de la intensidad y del tipo (positiva o negativa).

Sobre un eje de coordenadas se van marcando los puntos en que se cruzan los valores de la variable X y de la variable Y.

DISEÑO

Siguiendo a Kerlinger. "*"Diseño es el plan, estructura y estrategia de una investigación cuyo objetivo es dar respuesta a ciertas preguntas y controlar la varianza*".*

Como se aprecia, el autor define el término e incluye en su concepto los dos objetivos fundamentales del mismo.

DISEÑO EXPERIMENTAL

Sin entrar ahora en el segundo objetivo control* de la varianza*- el diseño en cuanto plan, estructura y estrategia nos permite poner a prueba, en condiciones rigurosas, la hipótesis del investigador sobre el efecto de una variable independiente sobre otra dependiente, contrastando el resultado con la posibilidad –probabilidad- de que tal efecto sea debido al azar.

DISTRIBUCIÓN DE FRECUENCIAS

En ocasiones, las puntuaciones originales o directas pueden representarse en una tabla en la que junto a estas –X_i- se incluya el número de veces que cada puntuación se repite (frecuencia: f_i).

Cuando el número de puntuaciones diferentes es elevado, las puntuaciones directas suelen sustituirse por conjunto de ellas, denominados *intervalos*; junto a ellos, las veces que ese conjunto de puntuaciones se repite (frecuencia del intervalo). En estos casos, a la hora de representar al intervalo (por ejemplo, para operar) se utiliza su valor medio, conocido como *marca de clase.*

ESCALAS DE MEDIDA

Al aplicar la regla de medida y la correspondiente unidad a un determinado objeto llegamos a un número. Pero los números resultantes no tienen todos las mismas propiedades ni, por tanto, se les pueden aplicar las mismas operaciones matemáticas.
Con los números más perfectos, propios de una *escala de medida de cociente o de razón* (edad, talla, peso) podemos utilizar todas las operaciones matemáticas. Con los de *escalas de intervalo* (temperatura), no, ya que no tienen un 0 absoluto. Hay números propios de una *escala de medida ordinal*, que admiten menos operaciones que los anteriores; ahora bien, dado que el *orden* tiene en alguna medida un carácter *cuantitativo* (por ejemplo: clase social) algunos autores clasifican, en ocasiones, a estas variables como *cuasi-cuantitativas*. Por último, hay números propios de escalas de medida nominal; aquí los números no indican cantidad sino diferencia: lo que es igual recibe el mismo número y lo que es diferente, un número distinto.

ESTADÍSTICA

Ciencia que trata de analizar e interpretar los datos recogidos con algún propósito, como la investigación científica.

Algunos autores la definen afirmando que su objeto es el estudio de los fenómenos aleatorios; recuerde el lector que cuando hablamos de contrastar los efectos de diversas intervenciones lo que hacemos es asignar probabilidades a que tales efectos se deban al puro azar (aleatoriedad) o a la intervención llevada a cabo por el investigador en condiciones de rigor o control* de explicaciones alternativas.

Cuando trabajamos con los valores de las muestras la Estadística se denomina *descriptiva;* si de tales valores deseamos pasar a estimar los correspondientes a la población, la Estadística se conoce como *inferencial*; esta es más compleja pero es la que ofrece más utilidad u aplicaciones tanto al científico como al profesional.

La inferencia estadística pretende sacar conclusiones sobre gran número de datos a través de observaciones de parte de esos datos. Se trata de generalizar los datos de una muestra a la población de la que procede. Mediante la estadística inferencial se puede estimar parámetros y realizar contraste de hipótesis.

ESTADÍSTICO

Valores obtenidos en una muestra. Los más conocidos son los agrupados bajo las medidas de posición o tendencia central (media, mediana*, moda*), las de dispersión* o variabilidad (desviación media, desviación típica, varianza) o los coeficientes de correlación*. Suelen representarse con letras latinas ($\overline{x}$, s, r…).

A partir de ellos, por inferencia estadística, podemos estimar sus correspondientes parámetros con determinados niveles de probabilidad, asumiendo un riesgo de error tipo I prefijado por el investigador.

ESTIMACIÓN DE PARÁMETROS

Se denomina así el procedimiento por el que se trata de estimar el valor de un estadístico, obtenido en una muestra, a toda la población de la que aquella forma parte.

Toda estimación asume un cierto margen de error, medido en términos de probabilidad; este error puede hacerse tan pequeño como desee el investigador, pero nunca podrá hablar en términos seguros, de certeza.

Al hablar en el texto del coeficiente de correlación nos hemos acercado al concepto y procedimiento de estimación de parámetros.

EXPERIMENTO

Es la modalidad de investigación empírica* más exigente; como consecuencia, su aportación esencial es la posibilidad de establecer, con razonable seguridad, relaciones de causa a efecto entre una o varias variables independientes (v.i.) y otra denominada dependiente (v.d.).

Para poder lograrlo se deben cumplir determinadas exigencias: el investigador debe poder planificar la acción y provocar el fenómeno, ha de poder realizarlo en condiciones de control y debe contar con medidas de calidad, tan válidas, fiables y precisas como sea posible.

FIABILIDAD

La fiabilidad es una de las características técnicas que deben reunir los instrumentos de recogida de datos y de medida. Entendemos por fiabilidad el grado de precisión de la medida; a más precisión –fiabilidad- menos error de medida.

La fiabilidad se expresa en muchos casos mediante un coeficiente de correlación (representado aquí por r_{XX}, dado que las dos series tienen que ver con el mismo instrumento de recogida de datos).

Las dos series pueden referirse a las puntuaciones en dos mitades de la misma prueba (*consistencia interna*), a las puntuaciones de la misma prueba aplica en dos ocasiones separadas por un cierto tiempo (*estabilidad*) o a las puntuaciones en dos pruebas equivalentes, esto es, de características tan semejantes como sea posible (*equivalencia*)

GENERALIZACIÓN

Entendemos por generalización al hecho de extender los resultados de la investigación desde los sujetos estudiados a los grupos de los que forma parte, desde las muestras a sus respectivas poblaciones.

Este procedimiento, propio de la Estadística inferencial, tiene exigencias que deben cumplirse para que sea legítimo y correcto.

HIPÓTESIS

Entendemos por hipótesis las conjeturas sobre la posible relación entre los elementos -variables- integrantes del problema. En los diseños experimentales se formulan hipótesis sobre la relación causal entre una o varias variables independientes (V.I.) y la variable dependiente (V.D.)

Una hipótesis se somete a prueba o se contrasta tratando de apreciar si las probabilidades a su favor son sensiblemente superiores a una explicación por azar. Esta segunda hipótesis se denomina nula y se representa por H_0 , frente a la del investigador (H_1)

HISTOGRAMA

Representación gráfica de las puntuaciones obtenidas por un conjunto de sujetos en una variable cuantitativa. En el eje X se sitúan los límites de los intervalos; en el Y, la frecuencia del intervalo.

INVESTIGACIÓN EMPÍRICA

Para Selltiz, *"Investigar es buscar de nuevo, echar otra mirada más cuidadosa para averiguar más. Echamos otra mirada porque puede haber algo erróneo en lo que ya sabemos [...]*

La investigación científica ha de ser *sistemática, organizada, disciplinada y rigurosa.*

Investigación empírica es aquella que acude a la experiencia, a los datos, para llegar a conclusiones en relación con las hipótesis de partida.

MEDIA ARITMÉTICA

Medida de posición resultante de sumar todas las puntuaciones de un grupo y dividir el resultado por el número de integrantes del grupo, representado por N.

Su ventaja fundamental radica en que todas y cada una de las puntuaciones de la serie incluyen en su valor en forma proporcional al mismo. Es especialmente adecuada para niveles de medida de razón e intervalo.

MEDIANA

Medida de posición resultante de ordenar las puntuaciones de mayor o menor, o viceversa, y encontrar la que ocupa el lugar central de la serie. Si la serie tiene un número par de casos, la mediana será la media de las dos centrales.

Su inconveniente fundamental es que en la mediana no influyen los valores de las puntuaciones sino solo el orden que ocupan. Dos series muy diferentes pueden tener la misma mediana.

Resulta especialmente adecuada para el nivel de medida ordinal.

MEDIDA. ESCALAS DE MEDIDA

Una medida, en sentido estricto, es el resultado de comparar una unidad con una cantidad. La cantidad "peso" la medimos comparándola con la unidad "Kilogramo" u otras mayores o menores. El resultado es el número.

La definición más amplia de "medida" se debe a Stevens: Medir es asignar numerales a los objetos o hechos de acuerdo con ciertas reglas. Un numeral puede ser un número o un símbolo, lo que permite admitir el nivel o escala de medida nominal.

En nuestros ámbitos, no siempre es tan fácil proceder a medir variables; la mayoría de las variables son construcciones o constructos elaborados por los científicos e investigadores, como en el caso de la inteligencia, el nivel de conocimientos, el autoconcepto, la tasa de inflación, el producto interior bruto o similares.

En tales casos, la medida consiste en la asignación de valores de acuerdo con ciertas reglas, como ocurre en una prueba objetiva, un cuestionario de actitudes hacia los inmigrantes, la tasa de mortalidad infantil, etc. Los números que resultan no tienen las mismas propiedades que en el caso del peso, de la talla o de la edad, números perfectos que permiten todo tipo de operaciones y que son propios de escalas de medida de *razón o cociente*.

Variables como la temperatura, perfectamente medibles, se diferencian de las anteriores en que el punto de partida –cero grados- no es fijo, además de poder presentar valores inferiores. Este tipo de variables forman parte de la escala de *intervalos*. Las que se limitan a indicar el *orden* en una serie (primero, segundo…) se ubican en las escalas *ordinales*; y en el caso de variables que no indican cantidad sino semejanza o diferencia (sexo, estado civil, clase social, grados universitarios…) la escala se conoce como *nominal.*

MODA

También denominada Modo, es una medida de posición que coincide con el valor más repetido de la serie de valores.

Su inconveniente fundamental es que en aquellos valores menos repetidos que el de la Moda no cuentan para su obtención.

Resulta especialmente adecuada para el nivel de medida nominal.

MODELO

Entendemos por "modelo" una representación simplificada de la realidad. Tal representación puede ser icónica, analógica, matemática.

Los modelos matemáticos tienen una gran utilidad en Estadística. En la medida en que unos datos empíricos sigan razonablemente un modelo, podemos aplicar las propiedades de este al tratamiento estadístico de aquellos.

En nuestro ámbito, modelo es, un tipo de distribución de datos *teórico o ideal* al que pueden tender distribuciones *empíricas o reales* de ciertas variables.

Por ejemplo: la variable motivación por los idiomas, una vez medida en un conjunto amplio de sujetos (muestra) puede acercarse o apartarse más o menos de un modelo ideal o teórico como es la denominada curva normal de probabilidades* o campana de Gauss.

Este modelo tiene unas propiedades; si nuestros datos medidos se acercan suficientemente al modelo, podemos aplicarles las propiedades del mismo, lo que nos permitirá analizar los datos y obtener conclusiones.

Para decidir si podemos considerar que unos datos empíricos se acercan suficientemente al modelo hasta hacerlos compatibles con él, disponemos de pruebas de bondad de ajuste, como es el caso de chi o ji cuadrado, cuyo símbolo es χ^2.

Este tipo de pruebas asignan una probabilidad a los datos empíricos sobre su acomodación o no al modelo, lo que permite al investigador aceptar o no la hipótesis de nulidad.

MUESTRA. MUESTREO

Entendemos por *muestra* un subconjunto de una *población*. La muestra debe ser *representativa* de la población, para lo que deberá contar con un *tamaño suficiente* y con una selección por *procedimientos imparciales*, como el muestreo aleatorio.

Muestreo es el procedimiento utilizado para seleccionar la muestra; el preferible es el denominado *aleatorio simple.*

PARÁMETRO

Entendemos por parámetro el valor de un determinado estadístico no en la muestra en que se obtiene sino en el total de la población. Si los estadísticos más comunes, como las medidas de posición y variabilidad (media: $\overline{X}$; mediana: Md; moda*: Mo; desviación típica: s; varianza: s^2...) se suelen representar por letras latinas, los parámetros lo hacen por letras griegas (μ = media; σ = desviación típica; σ^2 = varianza...)

POBLACIÓN

El término "población" se define como el conjunto de todos los casos o elementos que cumplen con las características que la definen: los varones, las mujeres, los estudiantes de Farmacia, los políticos, los abogados...

En ciencias sociales no suele estar muy claramente definida. El investigador desea generalizar los datos de la muestra a la población.

En los estudios empíricos no suele ser posible –ni, en la mayoría de los casos, aconsejable- estudiar todos los casos; se acude en su lugar a *muestras*, que deben ser representativas del conjunto total o población.

Por medio de la Estadística inferencial se pueden hacer estimaciones de los parámetros a partir de las muestras (por ejemplo: desde $\overline{X}$ a μ)

PROBABILIDAD

Frente a los sucesos seguros se encuentran los probables. El tipo de seguros a las que es más adecuado aplicar la probabilidad es el de los fenómenos aleatorios.

Conociendo las diferentes manifestaciones de un fenómeno, como el número de caras de un dado o de los números de la lotería, podemos decidir la denominada probabilidad *a priori,* suponiendo, como debe ocurrir, que todas las caras del dado y todos los números tienen las mismas oportunidades. En el primer caso, la probabilidad de una cara cualquiera es de 1/6; en el segundo, suponiendo que tengamos 60.000 números, será de 1/60.000.

Para nosotros es importante conocer los modelos de probabilidad, como el de la curva normal*. Gracias a ella, a la regla matemática que la rige, podemos asignar probabilidades a los fenómenos que la siguen, que se acomodan a ella. Estas probabilidades nos permiten aceptar o rechazar hipótesis (pruebas estadísticas) o decidir si un coeficiente de correlación* r_{XY} es o no estadísticamente significativo.

Los valores de probabilidad oscilan entre 0 –suceso imposible- y 1, suceso seguro.

Reflexione sobre la corrección o incorrección de una expresión habitual en los medios de comunicación como la siguiente: “casi con toda probabilidad…”. ¿Sería más correcto afirmar: con elevada probabilidad? ¿o, casi con seguridad?

PRUEBAS ESTADÍSTICAS. BONDAD DE AJUSTE

Cuando queremos contrastar dos o más formas de intervención –métodos, procedimiento de disciplina, sistemas de motivación, tratamientos fisioterapéuticos, fármacos…- acudimos a un contraste de hipótesis.

El investigador formula la suya –H_1- y, para darla por buena, debe ser capaz de rechazar la alternativa, conocida como hipótesis nula o de nulidad (H_0). Por prudencia, solo rechazará esta y aceptará aquella cuando las probabilidades a favor de esta sean tan pequeñas como desee y, en consecuencia, las probabilidades a favor de su hipótesis sean tan elevadas como él decida.

Las pruebas estadísticas más comunes son t y F, aunque hay otras muchas.

Un tipo de pruebas concreto es el de *bondad de ajuste*; se puede aplicar, por ejemplo, para decidir si una determinada distribución de datos se acomoda suficientemente a un modelo como para poder aplicar a aquellos las propiedades de este. Una de las más conocidas es la de χ^2 –chi o ji cuadrado) para decidir sobre la compatibilidad de una distribución de puntuaciones empíricas con la curva normal de probabilidades. Debemos decir que hay más pruebas y más modelos.

PUNTUACIÓN

Valor, generalmente numérico, que se atribuye a cada una de las manifestaciones de una variable.

La puntuación directa o bruta de un sujeto en una variables se representa por X_i (puntuación directa del sujeto i en la variable X). Cuando esta puntuación se pone en relación con la media

del grupo (X_i – Media) estamos ante una puntuación diferencial (x_i). Si esta puntuación diferencial se divide por la desviación típica del grupo estamos ante la puntuación z_i, que indica en cuantas desviaciones típicas se aparta un sujeto de la media del grupo.

Cuando afirmamos que Guadalajara se aparte de Madrid 55 KM. estamos diciendo algo similar a cuando decimos que la z de un sujeto es -2 (se aparta de la media dos veces la desviación típica, como Guadalajara se aparta 55 veces el valor de un Km.

Otro tipo de puntuación es el *cuantil,* una medida que interpreta las puntuaciones directas ordenadas, divididas en 4, 10 o 100 partes (cuartiles: Q; deciles: D; o centiles o percentiles: C o P). Si ese grupo muestral se considera representativo de la correspondiente población, podemos interpretar las puntuaciones utilizándolo como baremo o regla de medida. Así podemos afirmar que un sujeto está en el Q_2, en el D_6 o en el P_{81}, esto es, entre el 25 y el 50 % de los caso, en el 60 % superior o dejando por debajo de sí 81 casos de cada 100.

SIGNIFICACIÓN ESTADÍSTICA

Por lo general, todo investigador está interesado en saber si los valores obtenidos en una muestra, denominados *estadísticos,* representan a los de toda la población (*parámetros).*

A este procedimiento lo hemos denominado *estimación de parámetros.* Cuando el valor medido en una muestra representa al valor para toda la población afirmamos que ese estadístico es estadísticamente significativo. Si no fuera así, no podríamos considerar al citado estadístico como representante del parámetro: parámetro y estadístico serian valores de poblaciones diferentes.

Como hemos señalado, toda estimación asume un cierto margen de error, medido en términos de probabilidad; este error puede hacerse tan pequeño como desee el investigador, pero nunca podrá hablar en términos seguros, de certeza.

Algo similar podemos afirmar en los contrastes de hipótesis. Cuando un investigado plantea su hipótesis, por ejemplo: los resultados sobre el clima de aula –variable dependiente- serán mejores con un sistema A de disciplina que con otro B –variable independiente- (H_1) trata de mantener su hipótesis frente a una hipótesis alternativa –hipótesis nula o de nulidad, H_0-

Al final, después de aplica durante un tiempo los dos sistemas, llegará, por ejemplo, a dos medias aritméticas, y su problema será el de decidir si la diferencia entre ambas puede atribuirse a que el sistema A es mejor que el B o puede explicarse por casualidad, por azar (H0).

Si puede hacer lo primero, afirmará que las diferencias entre ambas medias aritméticas son reales, son estadísticamente significativas, y podrá mantener H_1 con una probabilidad a su favor tan elevada como desee, pero nunca con certeza. En caso contrario, no podrá rechazar H_0 y tendrá que admitir que tales diferencias pueden ser explicadas por el azar.

VALIDEZ

Utilizamos el término “validez” en dos contextos diferentes:

a) Como cualidad técnica de un instrumento de recogida de datos, indicando el grado en que tal instrumento mide lo que pretende y dice medir.

Como hemos indicado en el texto, dos manifestaciones de la validez, la *concurrente* y la *predictiva*, utilizan la correlación para poner de relieve la magnitud de la misma.

b) Como exigencia fundamental en los diseños de investigación experimental. La denominada *validez interna*, de darse, permite afirmar que los efectos medidos en la variable dependiente se deben a, y solo a, la variable independiente. Para ello el investigador debe controlar las variables extrañas. La *validez externa* se conoce como generalización, e informa del grado en que los resultados de la investigación pueden generalizarse.

VARIABLES

Frente a una *constante*, la variable es aquella realidad que admite diversos valores, como la edad, la clase social, la inteligencia, el rendimiento académico o diferentes dimensiones o factores de la personalidad.

Cuando una variable solo admite valores enteros la denominamos *discreta*, tal como ocurre con el sexo, el estado civil, la clase social, o la carrera universitaria; las variables *continuas* pueden tener todo tipo de valores intermedios, como ocurre con la talla, el peso o la edad.

Las primeras pueden ser *dicotómicas,* si únicamente admiten dos valores o *politómicas*, en el caso contrario; en el primer caso se ha venido situando el *sexo*, mientras en el segundo podemos citar el *estado civil.*

Desde la perspectiva de la investigación las variables suelen clasificarse, en función del papel que desempeñan, en *independientes*, las manipuladas por el investigador, y *dependientes*, aquellas sobre las que se mide la influencia de las primeras; también podemos hablar de variables *extrañas*, esto es, variables que pueden convertirse en rivales de la hipótesis del investigador al influir sobre la dependiente junto a la independiente o en lugar de ella.

VARIANZA

Medida de *dispersión** o *variabilidad.* Estadísticamente es la media de la suma de las desviaciones individuales de un grupo de sujetos, elevadas al cuadrado, con respecto a la media aritmética de un grupo.

La varianza es un índice del grado en que las puntuaciones individuales se agrupan más o menos en torno a la media del grupo; si todas las puntuaciones individuales coincidieran con la media, la varianza sería 0; cuanto más se aparten de ella, mayor valor alcanzará la varianza.

Esta medida es muy importante en la Estadística inferencial ya que se utiliza en las pruebas de contraste de hipótesis; la más conocida e importante es la denominada F, o ANAVA (análisis de la varianza, aunque lo que se contrasta son medias aritméticas) que atribuye a las diferencias entre medias una determinada probabilidad de que no sean explicables como consecuencia del azar. En muchos textos encontrará la expresión ANOVA (de *analisys of variance*)

ANEXO. TABLA DE ÁREAS DE LA CURVA NORMAL DE PROBABILIDADES Y SU MANEJO

La tabla que se inserta a continuación presenta cinco columnas. La primera de ellas [(z, Puntuación tipificada (*x/σ*)] nos permite entrar por los diferentes valores de z_i correspondientes a las puntuaciones directas. Situados sobre el valor que nos interesa, encontramos a su derecha cuatro columnas:

- *Área desde la media a x/σ*
- *Área de la parte mayor*
- *Área de la parte menor*
- *y. Ordenada en x/σ*

Dejando de lado esta última columna, vamos ver cómo podemos manejar las otras tres.

a) Los valores a la derecha de nuestro valor z_i son valores en términos de probabilidad. Sabemos que la probabilidad oscila entre 0, suceso imposible, y 1, suceso seguro. Estos valores, multiplicados por 100, se convierten en %. Y para convertir esos valores de probabilidad o de % en número de casos, debemos hacer una simple regla de tres, teniendo en cuenta, en nuestro caso, el valor de N.

b) Si deseamos saber, en una serie de N = 78, suponiendo una distribución compatible con la normal, cuántos sujetos se sitúan entre z_i = -1 y la media, vamos a su derecha y encontramos 0.3413. Este valor de probabilidad equivale a un % de 34,13 que, para N = 78, supone 26,62 casos (27, por aproximación).

c) Si en esa misma serie deseamos saber cuánto casos quedan por encima, leemos en la columna siguiente *Área de la parte mayor* 0.8413, esto es, el 84,13 %, que representa 65,62 (66 por aproximación). Tenga en cuenta que al caso b) le estamos sumando el 50 % por encima de la media

d) Si nuestro interés fuera el de conocer a cuántos sujetos supera quien tiene z_i = -1, miraríamos en la columna siguiente, bajo el *Área de la parte menor*; vemos 0.1587, esto es, el 15,87 %, que, en número de casos es 12,38 (por aproximación, 12)

e) ¿Por qué hemos mirado el área de la parte menor o mayor? Basta con hacer un sencillo dibujo, situar z_i = -1 y ver si lo que buscamos es la probabilidad o el porcentaje menor, mayor o hasta la media. Le recomendamos que haga lo mismo con el valor de z_i positivo (z_i = 1)

f) Un caso diferentes puede ser aquel en el que buscamos valores de porcentajes o casos entre dos valores de z determinados y que estos estén ambos por encima o por debajo de la media, o uno por encima y otro por debajo. Veamos (recuerde la conveniencia de hacer el dibujo para su apreciación intuitiva):

   – Número de casos entre z_i = 1,2 y z_i = 2,1.
   En este caso buscaríamos en las tablas la probabilidad y % correspondiente a 2,1 hasta la media; haríamos lo mismo con 1,2 hasta la media y restaríamos: 0.4821 - 0.1151= 0.3670; 36,70 %; 28,62 casos (29, por aproximación)

– Número de casos entre z_i = -1,2 y z_i = -2,1.
  Debemos tener en cuenta que se trata de valores negativos, pero que la curva es simétrica. Por tanto, el mismo caso anterior.

– Número de casos entre z_i = -1,2 y z_i = 2,1.
  En esta ocasión, al estar un valor por encima de la media y otro por debajo de la misma, los valores de probabilidad y su traducción a % deben ser sumados. Por tanto: 0.4821 + 0.1151 = 0.5972; 59,72 %; 46,58 casos (47 por aproximación).

g) Una aplicación, coherente con todo lo anterior, es el caso inverso: ¿qué valores de z_i corresponden a determinados valores de probabilidad o de % dados? En estos casos, se debe entrar por la columna correspondiente (dos, tres o cuatro) y, una vez encontrado el valor o el más próximo, desplazarnos hacia la izquierda y leer el valor *x/σ* correspondiente. Así, encontrar el valor de z_i que deja por debajo de sí el 77,04 % supone entrar por la columna de *Área de la parte mayor* (obviamente, el 77,04 es la parte mayor) y, una vez encontrado, desplazarnos a la izquierda hasta encontrar en la primera columna el valor 0,74. Por tanto: zi = 0,74

h) Dado lo anterior, compruebe lo fácil que será calcular que puntuaciones z_i corresponderán a los diferentes cuantiles: cuartiles, deciles o percentiles. Eso sí, podrá ver en la tabla que, salvo para el Q_2 o P_{50} no encontrará valores exactos, por lo que entrará por los valores más próximos.

   Como Q_2 o P_{50} representan el 50 %, obviamente su z_i = 0

   Practiquemos con los siguientes valores: Q_3, D_6 y P_{81}: Los valores más próximo en las tablas, todos en *Área de la parte mayor,* son, en puntuaciones z_i (columna de la izquierda): 0,67, 0,25 y 0.88.

Tabla I.—Areas y ordenadas de la curva de distribución normal en función x/σ

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
0.00	.0000	.5000	.5000	.3989
0.01	.0040	.5040	.4960	.3989
0.02	.0080	.5080	.4920	.3989
0.03	.0120	.5120	.4880	.3988
0.04	.0160	.5160	.4840	.3986
0.05	.0199	.5199	.4801	.3984
0.06	.0239	.5239	.4761	.3982
0.07	.0279	.5279	.4721	.3980
0.08	.0319	.5319	.4681	.3977
0.09	.0359	.5359	.4641	.3973
0.10	.0398	.5398	.4602	.3970
0.11	.0438	.5438	.4562	.3965
0.12	.0478	.5478	.4522	.3961
0.13	.0517	.5517	.4483	.3956
0.14	.0557	.5557	.4443	.3951
0.15	.0596	.5596	.4404	.3945
0.16	.0636	.5636	.4364	.3939
0.17	.0675	.5675	.4325	.3932
0.18	.0714	.5714	.4286	.3925
0.19	.0753	.5753	.4247	.3918
0.20	.0793	.5793	.4207	.3910
0.21	.0832	.5832	.4168	.3902
0.22	.0871	.5871	.4129	.3894
0.23	.0910	.5910	.4090	.3885
0.24	.0948	.5948	.4052	.3876
0.25	.0987	.5987	.4013	.3867
0.26	.1026	.6026	.3974	.3857
0.27	.1064	.6064	.3936	.3847
0.28	.1103	.6103	.3897	.3836
0.29	.1141	.6141	.3859	.3825
0.30	.1179	.6179	.3821	.3814
0.31	.1217	.6217	.3783	.3802
0.32	.1255	.6255	.3745	.3790
0.33	.1293	.6293	.3707	.3778
0.34	.1331	.6331	.3669	.3765
0.35	.1368	.6368	.3632	.3752
0.36	.1406	.6406	.3594	.3739
0.37	.1443	.6443	.3557	.3725
0.38	.1480	.6480	.3520	.3712
0.39	.1517	.6517	.3483	.3697
0.40	.1554	.6554	.3446	.3683
0.41	.1591	.6591	.3409	.3668
0.42	.1628	.6628	.3372	.3653
0.43	.1664	.6664	.3336	.3637
0.44	.1700	.6700	.3300	.3621

Tomada del *Formulario y tablas de Pedagogía Experimental.* UNED, 1983, p. 119-126 (ambas inclusive).

Tabla I *(continuación)*

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
0.45	.1736	.6736	.3264	.3605
0.46	.1772	.6772	.3228	.3589
0.47	.1808	.6808	.3192	.3572
0.48	.1844	.6844	.3156	.3555
0.49	.1879	.6879	.3121	.3538
0.50	.1915	.6915	.3085	.3521
0.51	.1950	.6950	.3050	.3503
0.52	.1985	.6985	.3015	.3485
0.53	.2019	.7019	.2981	.3467
0.54	.2054	.7054	.2946	.3448
0.55	.2088	.7088	.2912	.3429
0.56	.2123	.7123	.2877	.3410
0.57	.2157	.7157	.2843	.3391
0.58	.2190	.7190	.2810	.3372
0.59	.2224	.7224	.2776	.3352
0.60	.2257	.7257	.2743	.3332
0.61	.2291	.7291	.2709	.3312
0.62	.2324	.7324	.2676	.3292
0.63	.2357	.7357	.2643	.3271
0.64	.2389	.7389	.2611	.3251
0.65	.2422	.7422	.2578	.3230
0.66	.2454	.7454	.2546	.3209
0.67	.2486	.7486	.2514	.3187
0.68	.2517	.7517	.2483	.3166
0.69	.2549	.7549	.2451	.3144
0.70	.2580	.7580	.2420	.3123
0.71	.2611	.7611	.2389	.3101
0.72	.2642	.7642	.2358	.3079
0.73	.2673	.7673	.2327	.3056
0.74	.2704	.7704	.2296	.3034
0.75	.2734	.7734	.2266	.3011
0.76	.2764	.7764	.2236	.2989
0.77	.2794	.7794	.2206	.2966
0.78	.2823	.7823	.2177	.2943
0.79	.2852	.7852	.2148	.2920
0.80	.2881	.7881	.2119	.2897
0.81	.2910	.7910	.2090	.2874
0.82	.2939	.7939	.2061	.2850
0.83	.2967	.7967	.2033	.2827
0.84	.2995	.7995	.2005	.2803
0.85	.3023	.8023	.1977	.2780
0.86	.3051	.8051	.1949	.2756
0.87	.3078	.8078	.1922	.2732
0.88	.3106	.8106	.1894	.2709
0.89	.3133	.8133	.1867	.2685

Tabla I *(continuación)*

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
0.90	.3159	.8159	.1841	.2661
0.91	.3186	.8186	.1814	.2637
0.92	.3212	.8212	.1788	.2613
0.93	.3238	.8238	.1762	.2589
0.94	.3264	.8264	.1736	.2565
0.95	.3289	.8289	.1711	.2541
0.96	.3315	.8315	.1685	.2516
0.97	.3340	.8340	.1660	.2492
0.98	.3365	.8365	.1635	.2468
0.99	.3389	.8389	.1611	.2444
1.00	.3413	.8413	.1587	.2420
1.01	.3438	.8438	.1562	.2396
1.02	.3461	.8461	.1539	.2371
1.03	.3485	.8485	.1515	.2347
1.04	.3508	.8508	.1492	.2323
1.05	.3531	.8531	.1469	.2299
1.06	.3554	.8554	.1446	.2275
1.07	.3577	.8577	.1423	.2251
1.08	.3599	.8599	.1401	.2227
1.09	.3621	.8621	.1379	.2203
1.10	.3643	.8643	.1357	.2179
1.11	.3665	.8665	.1335	.2155
1.12	.3686	.8686	.1314	.2131
1.13	.3708	.8708	.1292	.2107
1.14	.3729	.8729	.1271	.2083
1.15	.3749	.8749	.1251	.2059
1.16	.3770	.8770	.1230	.2036
1.17	.3790	.8790	.1210	.2012
1.18	.3810	.8810	.1190	.1989
1.19	.3830	.8830	.1170	.1965
1.20	.3849	.8849	.1151	.1942
1.21	.3869	.8869	.1131	.1919
1.22	.3888	.8888	.1112	.1895
1.23	.3907	.8907	.1093	.1872
1.24	.3925	.8925	.1075	.1849
1.25	.3944	.8944	.1056	.1826
1.26	.3962	.8962	.1038	.1804
1.27	.3980	.8980	.1020	.1781
1.28	.3997	.8997	.1003	.1758
1.29	.4015	.9015	.0985	.1736
1.30	.4032	.9032	.0968	.1714
1.31	.4049	.9049	.0951	.1691
1.32	.4066	.9066	.0934	.1669
1.33	.4082	.9082	.0918	.1647
1.34	.4099	.9099	.0901	.1626

Tabla I *(continuación)*

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
1.35	.4115	.9115	.0885	.1604
1.36	.4131	.9131	.0869	.1582
1.37	.4147	.9147	.0853	.1561
1.38	.4162	.9162	.0838	.1539
1.39	.4177	.9177	.0823	.1518
1.40	.4192	.9192	.0808	.1497
1.41	.4207	.9207	.0793	.1476
1.42	.4222	.9222	.0778	.1456
1.43	.4236	.9236	.0764	.1435
1.44	.4251	.9251	.0749	.1415
1.45	.4265	.9265	.0735	.1394
1.46	.4279	.9279	.0721	.1374
1.47	.4292	.9292	.0708	.1354
1.48	.4306	.9306	.0694	.1334
1.49	.4319	.9319	.0681	.1315
1.50	.4332	.9332	.0668	.1295
1.51	.4345	.9345	.0655	.1276
1.52	.4357	.9357	.0643	.1257
1.53	.4370	9370	.0630	.1238
1.54	.4382	.9382	.0618	.1219
1.55	.4394	.9394	.0606	.1200
1.56	.4406	.9406	.0594	.1182
1.57	.4418	.9418	.0582	.1163
1.58	.4429	.9429	.0571	.1145
1.59	.4441	.9441	.0559	.1127
1.60	.4452	.9452	.0548	.1109
1.61	.4463	.9463	.0537	.1092
1.62	.4474	.9474	.0526	.1074
1.63	.4484	.9484	.0516	.1057
1.64	.4495	.9495	.0505	.1040
1.65	.4505	.9505	.0495	.1023
1.66	.4515	.9515	.0485	.1006
1.67	.4525	.9525	.0475	.0989
1.68	.4535	.9535	.0465	.0973
1.69	.4545	.9545	.0455	.0957
1.70	.4554	.9554	.0446	.0940
1.71	.4564	.9564	.0436	.0925
1.72	.4573	.9573	.0427	.0909
1.73	.4582	.9582	.0418	.0893
1.74	.4591	.9591	.0409	.0878
1.75	.4599	.9599	.0401	.0863
1.76	.4608	.9608	.0392	.0848
1.77	.4616	.9616	.0384	.0833
1.78	.4625	.9625	.0375	.0818
1.79	.4633	.9633	.0367	.0804

TABLAS

Tabla I *(continuación)*

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
1.80	.4641	.9641	.0359	.0790
1.81	.4649	.9649	.0351	.0775
1.82	.4656	.9656	.0344	.0761
1.83	.4664	.9664	.0336	.0748
1.84	.4671	.9671	.0329	.0734
1.85	.4648	.9678	.0322	.0721
1.86	.4686	.9686	.0314	.0707
1.87	.4693	.9693	.0307	.0694
1.88	.4699	.9699	.0301	.0681
1.89	.4706	.9706	.0294	.0669
1.90	.4713	.9713	.0287	.0656
1.91	.4719	.9719	.0281	.0644
1.92	.4726	.9726	.0274	.0632
1.93	.4732	.9732	.0268	.0620
1.94	.4738	.9738	.0262	.0608
1.95	.4744	.9744	.0256	.0596
1.96	.4750	.9750	.0250	.0584
1.97	.4756	.9756	.0244	.0573
1.98	.4761	.9761	.0239	.0562
1.99	.4767	.9767	.0233	.0551
2.00	.4772	.9772	.0228	.0540
2.01	.4778	.9778	.0222	.0529
2.02	.4783	.9783	.0217	.0519
2.03	.4788	.9788	.0212	.0508
2.04	.4793	.9793	.0207	.0498
2.05	.4798	.9798	.0202	.0488
2.06	.4803	.9803	.0197	.0478
2.07	.4808	.9808	.0192	.0468
2.08	.4812	.9812	.0188	.0459
2.09	.4817	.9817	.0183	.0449
2.10	.4821	.9821	.0179	.0440
2.11	.4826	.9826	.0174	.0431
2.12	.4830	.9830	.0170	.0422
2.13	.4834	.9834	.0166	.0413
2.14	.4838	.9838	.0162	.0404
2.15	.4842	.9842	.0158	.0396
2.16	.4846	.9846	.0154	.0387
2.17	.4850	.9850	.0150	.0379
2.18	.4854	.9854	.0146	.0371
2.19	.4857	.9857	.0143	.0363
2.20	.4861	.9861	.0139	.0355
2.21	.4864	.9864	.0136	.0347
2.22	.4868	.9868	.0132	.0339
2.23	.4871	.9871	.0129	.0332
2.24	.4875	.9875	.0125	.0325

Tabla I *(continuación)*

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
2.25	.4878	.9878	.0122	.0317
2.26	.4881	.9881	.0119	.0310
2.27	.4884	.9884	.0116	.0303
2.28	.4887	.9887	.0113	.0297
2.29	.4890	.9890	.0110	.0290
2.30	.4893	.9893	.0107	.0283
2.31	.4896	.9896	.0104	.0277
2.32	.4898	.9898	.0102	.0270
2.33	.4901	.9901	.0099	.0264
2.34	.4904	.9904	.0096	.0258
2.35	.4906	.9906	.0094	.0252
2.36	.4909	.9909	.0091	.0246
2.37	.4911	.9911	.0089	.0241
2.38	.4913	.9913	.0087	.0235
2.39	.4916	.9916	.0084	.0229
2.40	.4918	.9918	.0082	.0224
2.41	.4920	.9920	.0080	.0219
2.42	.4922	.9922	.0078	.0213
2.43	.4925	.9925	.0075	.0208
2.44	.4927	.9927	.0073	.0203
2.45	.4929	.9929	.0071	.0198
2.46	.4931	.9931	.0069	.0194
2.47	.4932	.9932	.0068	.0189
2.48	.4934	.9934	.0066	.0184
2.49	.4936	.9936	.0064	.0180
2.50	.4938	.9938	.0062	.0175
2.51	.4940	.9940	.0060	.0171
2.52	.4941	.9941	.0059	.0167
2.53	.4943	.9943	.0057	.0163
2.54	.4945	.9945	.0055	.0158
2.55	.4946	.9946	.0054	.0154
2.56	.4948	.9948	.0052	.0151
2.57	.4949	.9949	.0051	.0147
2.58	.4951	.9951	.0049	.0143
2.59	.4952	.9952	.0048	.0139
2.60	.4953	.9953	.0047	.0136
2.61	.4955	.9955	.0045	.0132
2.62	.4956	.9956	.0044	.0129
2.63	.4957	.9957	.0043	.0126
2.64	.4959	.9959	.0041	.0122
2.65	.4960	.9960	.0040	.0119
2.66	.4961	.9961	.0039	.0116
2.67	.4962	.9962	.0038	.0113
2.68	.4963	.9963	.0037	.0110
2.69	.4964	.9964	.0036	.0107

Tabla I *(continuación)*

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
2.70	.4965	.9965	.0035	.0104
2.71	.4966	.9966	.0034	.0101
2.72	.4967	.9967	.0033	.0099
2.73	.4968	.9968	.0032	.0096
2.74	.4969	.9969	.0031	.0093
2.75	.4970	.9970	.0030	.0091
2.76	.4971	.9971	.0029	.0088
2.77	.4972	.9972	.0028	.0086
2.78	.4973	.9973	.0027	.0084
2.79	.4974	.9974	.0026	.0081
2.80	.4974	.9974	.0026	.0079
2.81	.4975	.9975	.0025	.0077
2.82	.4976	.9976	.0024	.0075
2.83	.4977	.9977	.0023	.0073
2.84	.4977	.9977	.0023	.0071
2.85	.4978	.9978	.0022	.0069
2.86	.4979	.9979	.0021	.0067
2.87	.4979	.9979	.0021	.0065
2.88	.4980	.9980	.0020	.0063
2.89	.4981	.9981	.0019	.0061
2.90	.4981	.9981	.0019	.0060
2.91	.4982	.9982	.0018	.0058
2.92	.4982	.9982	.0018	.0056
2.93	.4983	.9983	.0017	.0055
2.94	.4984	.9984	.0016	.0053
2.95	.4984	.9984	.0016	.0051
2.96	.4985	.9985	.0015	.0050
2.97	.4985	.9985	.0015	.0048
2.98	.4986	.9986	.0014	.0047
2.99	.4986	.9986	.0014	.0046
3.00	.4987	.9987	.0013	.0044
3.01	.4987	.9987	.0013	.0043
3.02	.4987	.9987	.0013	.0042
3.03	.4988	.9988	.0012	.0040
3.04	.4988	.9988	.0012	.0039
3.05	.4989	.9989	.0011	.0038
3.06	.4989	.9989	.0011	.0037
3.07	.4989	.9989	.0011	.0036
3.08	.4990	.9990	.0010	.0035
3.09	.4990	.9990	.0010	.0034
3.10	.4990	.9990	.0010	.0033
3.11	.4991	.9991	.0009	.0032
3.12	.4991	.9991	.0009	.0031
3.13	.4991	.9991	.0009	.0030
3.14	.4992	.9992	.0008	.0029

Tabla I *(continuación)*

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
3.15	.4992	.9992	.0008	.0028
3.16	.4992	.9992	.0008	.0027
3.17	.4992	.9992	.0008	.0026
3.18	.4993	.9993	.0007	.0025
3.19	.4993	.9993	.0007	.0025
3.20	.4993	.9993	.0007	.0024
3.21	.4993	.9993	.0007	.0023
3.22	.4994	.9994	.0006	.0022
3.23	.4994	.9994	.0006	.0022
3.24	.4994	.9994	.0006	.0021
3.30	.4995	.9995	.0005	.0017
3.40	.4997	.9997	.0003	.0012
3.50	.4998	.9998	.0002	.0009
3.60	.4998	.9998	.0002	.0006
3.70	.4999	.9999	.0001	.0004

Tomada de DOWNIE, N. M. y HEATH, R. W.: *Métodos Estadísticos Aplicados*. Ed. del Castillo, Madrid, 1979, p. 320-327.